VOL. XXXV
No. 2

PSYCHOLOGICAL REVIEW PUBLICATIONS

WHOLE NO. 163
1926

Psychological Monographs

Studies in Psychology

From the University of Illinois

EDITED BY

MADISON BENTLEY

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, N. J.
AND ALBANY, N. Y.

Agents: G. E. STECHERT & CO., London (2 Star Yard, Carey St., W. C.)
Paris (16 rue de Condé)

TABLE OF CONTENTS

THE RELATION OF "INSTRUCTION" TO THE PSYCHOSOMATIC FUNCTIONS

By Oscar F. Weber and Madison Bentley

The Problem

In sustaining its intercourse with nature, as well as for its own internal regulation, the human organism has recourse to a considerable number of psychosomatic or "mind-body" functions or modes of operation. These functions may be indicated by the terms perception, memory, imagination, action, emotion, comprehension and elaboration or thinking. Like all organic operations these functions proceed from definable conditions; that is to say, they are induced by specific antecedents, which serve to throw the organism into commission for a given operation. On the bodily side, are neural trends or sets, glandular reserves, disposable energy and other available means. On the mental side, or (to be more exact) within phenomenal experience, are to be found intents, anticipations, forward-looking hints, premonitions and commands. Under our common conditions in the laboratory the anticipation of function usually appears in the formal instruction, "Move quickly when the stimulus light appears," "Attend to the fixation mark," "Read the syllables," and so on. The same form of preparation is likewise given in common life by such phrases as "Look at the valley when we approach the next turn," "Hold your hat on" and "Lend me a match." But just as much, whether we are in the street or in the laboratory, does the *occasion itself* (*e.g.,* the wayside fountain, the approach of the street car, the half-familiar face, the fire-siren) or the *organism itself* ("I must get to work," "What *was* that name?" "I can if I try") "instruct" the individual toward some functional performance. Thus is the psychological organism constantly set, tuned, bullied, encouraged or charged for seeing and for hearing, for remembering, for acting and suffering emotion,

for understanding and for thinking. Always is the body prepared; and phenomenal experience also contains premonitions, prophecies, announcements and intimations of imminent and forthcoming tasks and operations.

And not only do these functional antecedents throw the organism into commission for its performances. Since they stand as compact surrogates for the animal's history, they also —if they fit the occasion—facilitate an approaching function, give it precision, inform it with significance, and render it both apt and effective.

Now the psychological experiment, with its trained observers and its nice control of conditions, as well as the less analytic test, have made it plain that such performances of the organism as we have indicated are carried through with high variability. Individuals vary in the accomplishment of one and the same task and the single individual himself varies from time to time and from occasion to occasion. These wide variations are generally laid to differences of a vaguely conceived "capacity," though at times the operation of diverse and multitudinous "mental functions" is also invoked. It is our intent to suggest in this study that the main factor to be taken into account, both in understanding the functional operations themselves and in supplying a key to the wide variability of their output or accomplishment, lies elsewhere; that it lies in the *antecedents of function* and in the *precise resources of the organism* which they throw into commission for the task in hand.

In order to make our study as empirical as we can, we have, in what follows, left aside the bodily factors of functional preparation, where our knowledge is, for the greater part, both hypothetical and vague; and we have confined ourselves to the experiential factors, to those antecedent factors which are at hand for the report of the observer and which depict the actual way in which the organism devotes itself to the discharge of an impending function. Taken all together we shall use the common term "instruction" to include all the experiential or phenomenal facts which lead up to and usher in the functional performance.

Among these factors we shall distinguish three types of instruction, "formal," "occasional" and "self-imposed," and we shall discuss the relation of instruction to attitude, posture, predisposition, task and Aufgabe.

Historical summary. Külpe's early insistence upon the determination of action by instruction (in the form of expectant attention) and the primary use by Ach, Watt, Messer and other representatives of the Würzburg School of the concept of *Aufgabe* in the initiation and progress of thinking are everywhere known in psychology.[1] Meumann[2] noted that the associative reaction depends upon the form of the instruction ("respond as quickly as possible," "as effectively as possible," etc.); Whipple[3] that the stimulus threshold for pain varies with the verbal instruction ("report when the pressure is mildly disagreeable," "is uncomfortable," "begins to hurt," etc.); Fernberger[4] that the hint to look for a "difference" affected the magnitude of the differential limen; George[5] made it clear that the "doubtful" judgment may rest upon an ambiguous or shifting instruction and has therefore to be carefully controlled in the psychophysical metric methods, and Godfrey Thomson[6] approached the concept of instruction and its influence upon function in his naïve comment that the mode of psychophysical report depends upon "a conscious act of the subject, and can be varied, if he is so disposed, at his whim." Many experimental researches have made explicit study of the effect in various fields of the form of instruction upon report. We may mention Friedländer, Fernberger and Reid with lifted weights,[7] Pratt[8] with tonal complexes, and Hoisington and others[9] with limens for film color. Occasionally some attention has been paid to the control of the instruction in the administering of the test;[10] albeit the setting of a gross task usually takes for granted there the actual instruction under which the subject labors. We have abundant evidence, however, that high training and careful regard of observational errors are necessary to insure functional performance under a fixed and unequivocal instruction. Mere docility and faithful intent to do as the administrator of a test generally directs offer no guarantee whatsoever either that the psychosomatic operations are called forth as the verbal instructions imply or that the same or even similar functional resources are appealed to by different observers.

[1] See, for literature and discussion, BENTLEY, M., *The Field of Psychology,* 1924, 360ff, 384ff, 452, 505, 533ff.

[2] MEUMANN, E. *Vorlesungen zur Einführung in die experimentelle Pädagogik,* 1913, Vol. II, 420.

[3] WHIPPLE, G. M. *Manual of Mental and Physical Tests,* 1914, Pt. I, 236.

[4] FERNBERGER, S. W. *Amer. J. of Psychol.,* 1914, Vol. XXV, 538–543.

[5] GEORGE, S. *Ibid.,* 1917, Vol. XXVIII, 1–37.

[6] THOMSON, G. *Psychol. Rev.,* 1920, Vol. XXVII, 300–307. Cf. Boring, E. G., *Ibid.,* 440–452.

[7] FRIEDLÄNDER, H. *Zsch. f. Psychol.,* 1920, Vol. LXXXIII, 129–210; FERNBERGER, S. W., *J. of Exper. Psychol.,* 1921, Vol. IV, 63–76; REID, A. C., *Amer. J. of Psychol.,* 1924, Vol. XXXV, 53–74.

[8] PRATT, C. C. *Ibid.,* 1921, Vol. XXXII, 490–515.

[9] ELLIOT, M., WEST, J., and HOISINGTON, L. B. *Ibid.,* 1924, Vol. XXXV, 125–131.

[10] JUDD, C. H., and BUSWELL, G. T. Silent Reading; a Study of various Types. Univ. of Chicago (*Educ. Monog. Suppl.,* No. 23), 1922.

The Experiments

The general procedure: Observers[11] were given, under verbal instruction, such simple tasks as the reading of plain sentences, of unspaced sentences, or of sentences with colons placed between succeeding letters. The Os were informed of the general nature of the study (the exact reporting of all forms and kinds of instruction appearing before and during the task) and they were warned that they could never safely assume that the verbal formula of the experimenter was the *actual* instruction under which they proceeded. The following description was supplied upon a sheet of paper to each observer before he began his observations, and it was subsequently kept near him for reference.

The Nature of "Instruction"

Any preparation of the organism for the performance of a mind-body function might be looked upon as an "instruction." Thus the organism might be said to be prepared or "instructed" to do so-and-so, *i.e.*, to observe, to understand, to remember, and the like. This use of the term "instruction" would, however, include such purely hypothetical factors as neural "trends," "sets" and "predispositions," as well as directly observable and reportable anticipations, commands, and forecasts. Since we are here concerned with matters of observation only, we shall neglect these hypothetical factors and consider under "instruction" only those forms of preparation which fall within the commentary of the observer or subject.

These forms are three, namely, (1) the formally imposed instruction, (2) the self-imposed instruction, and (3) the occasional instruction.

The *formally imposed instruction* is verbal. It is given by another individual, as in the instruction accompanying a test; *e.g.*, "In order to add the following numbers you may rearrange them," or "Can you find suitable verb-forms to fill in the blanks in the following sentences?" or "Study the chart on the next page." The formal instruction is, then, a verbal statement, query or command, oral or written, as it is comprehended by the subject.

The *self-imposed instruction* is the instruction which the individual gives himself, *e.g.*, "I must add as fast as I can" (command), "Shall I look at the whole picture first?" (query), "I may take my time" (comment), or "I must not be distracted" (negative command).

In the *occasional instruction* the subject is incited to his task by some feature of the task itself, which directly suggests the mode of performance or the end to be achieved; *e.g.*, the expression

$$\begin{array}{r} 454 \\ 5 \\ \hline \end{array}$$

[11] Professor Bentley (B), Dr. Griffith (G), Dr. Young (Y), Dr. Culler (C), Dr. Kingsley (K), Miss Anderson (A), and Miss Hopkins (Ho), members of the staff of the Department of Psychology, and Mr. Mikesell (M), Miss Wedekind (W), and Miss Hatfield (Ha), graduate students.

presented without words from the experimenter would be likely itself to give the instruction "Multiply," whereas a column of three-place numbers, as

621
278
342
534

would be likely to give the instruction "Add these numbers." But the occasion may be the source of instruction in other ways, *e.g.*, any of the factors, light, the time of day, or the tone of the experimenter, might conceivably thus serve.

You will find, furthermore, that each of these instructions takes two forms; namely, positive and negative, preparing either for or against a given type of performance. As you carry out the tasks to be assigned, note these forms and mention them as you find them.

The Os were familiar with the two forms of report, the *inspection* (a descriptive account of qualitative variety, including the intensive, durational and other degrees in which a given quality might appear) and the *commentary* (O's historical account of his experiences related in terms of performance and of the conditions of performance.[12] It was made plain that the main report was to be made in the form of the commentary. The tabular outline of Table I was supplied for each observation. Columns 2, 3 and 4 have been filled in from sample observations to make clear the character of the report. Column 2 gives the sequence of the instructions reported, column 3 the psychosomatic function which the instruction induced, and column 4 the functional product, in terms of things perceived, understood, hinted at, guessed, queried, thought out, etc. After a "ready" signal, E exposed the typed command "Read the following," and thereupon visually appeared for a few seconds and at a convenient reading distance the typed sentence,

FEW ARE ABLE TO GRANT TO OTHERS THE RIGHT TO ERR.

The results, put down in terms of instruction, of functions elicited, and of functional products are exemplified by the following Table.

[12] The terms employed and the conception of function follow Bentley, *The Field of Psychology*, 1924, 189ff, 384ff. For the distinction drawn between "inspection" and "commentary," as forms of the psychological report, see pp. 20, 41–44, 197–199, 208.

TABLE I

Form of instruction	Obs.	Rep'ted sequ'ce	P-S funct'ns introduced	Functional products[13]
Formal	B	1	C	E's words understood
	C	1	P & C	Sentence seen and generally understood
	Y	1	C	E's words understood
Self	B	2	C	I am to get the meaning (Vb)
	C	3	C	This is what I am expected to do (Vb)
	Y	2	C	Just what does it mean? (Ps)
		4		The sentence is ambiguous (Kn)
Occasional	B	3	C	The words challenge comprehension (Ps)
		4	El	That is a thought problem (Vb)
		5	E	Ah! that solution is satisfactory (Ps)
	C	2	M	Sentence suggested a passage in ——
		4	E	This task is foolish: I resent it
	Y	3	El	Thought problem (stage of formulation)
		5	El	Is the statement true? (Vb)

P = perception. M = memory. I = imagination. A = action. E = emotion. C = comprehension. El = elaboration or thinking. (Ps) = posture (Bewusstseinslage, attitude). (Vb) = verbalization. (K) = kinesthetic vehicle.

Results

Series I. Besides the sentence quoted above, the three following statements were employed under the same conditions:

ANATOLE FRANCE HAS PRONOUNCED H. G. WELLS TO BE THE GREATEST INTELLECTUAL FORCE IN THE WORLD TO-DAY.

THE PUPILS INSISTED THEY HAD THOUGHT TO RECITE TO BE TO READ THE EXERCISE.

WHAT WILL BE McA——'S FIRST MOVE IN CHICAGO?

To avoid the decay through repetition of the first formal (verbal) instruction the form of the phrase was constantly changed. Variations employed were "See what follows," "I want you to read the following" and "Can you read the following?" Even with these variations this part of the instruction tended to change into the occasional kind, in which the experimental setting anticipated the entire exposure, finally to be abbreviated into the perceptive-kinesthetic flash, common in the reaction experiment, with the meaning "Here comes the task."

When allowance is made for differences in completeness and accuracy from the various O's, we observe that the results are practically the same for all and for all four sentences. Beginning with the formal instruction (F-I), "Read, etc.," which led to a preliminary comprehension, the task was then sustained by from 2 to 5 self (S-I) and occasional (O-I) instructions which

[13] It is necessary to make our reports in verbal form. The reader must remember, however, that in many cases neither the instruction (self and occasional) nor the functional performance itself proceeded by way of words. Many postures (*Bewusstseinslagen*), tendinous strains, organic configurations, and so on, served to carry both the instruction and the operations themselves. At times a general attitude (as of willingness, skepticism, rivalry, curiosity) appeared before the experiment began and carried right through, exerting its strong influence upon the entire course of events. The "good" observer knows how to neutralize, or at the least to allow for, these fixed settings which often predetermine his results and which not infrequently make the reports of the "poor" observer absolutely worthless.

issued in comprehension and elaboration. Once or twice a memorial or imaginational apprehension appeared; once a distinct action came from the self-instruction, "You must sweep your eyes across"; and two or three emotive stirs cut across comprehension or elaboration. About three times as many S-I and O-I appeared as instructions of the formal sort. Had we included the preliminary period of adjustment (*Vorperiode*), doubtless we should have increased the ratio. Although the exposure-time was tempered to the reader's rate, it never exceeded 20 seconds.

Series II. To shift the interplay of instructions we now increased the difficulty of the task, (a) by sense sentences typed with equal spacing—

N O W I T H O U G H T T H E E N D H A D C O M E
I N E A S E A N D C O M F O R T H E S A T H E R E
O N A N D O N A R G U E D T H E A G E D A N G L E R
T O G O E T H E L A W A N D L I F E W E R E O N E

and (b) with the intervening colon—

O:F:A:L:L:M:E:N:H:E:W:A:S:L:E:A:S:T:W:E:L:C:O:M:E
I:N:A:N:I:N:A:N:E:W:A:Y:O:N:A:N:D:O:N:H:E:R:A:M:B:L:E:D
I:S:T:H:E:I:R:S:H:I:B:B:O:L:E:T:H:M:O:S:E:S:A:H:O:L:I:A:B:
A:N:D:B:E:Z:A:L:E:E:L:O:R:J:O:S:H:U:A:Z:E:R:U:B:B:A:B:
E:L:A:N:D:H:A:G:G:A:I:?

The literate and bookish adult is so constantly primed for print that a simple formal command to read might be assumed to be sufficient to touch off that facile and highly integrated train of perceptive-actional-comprehending functions which we know as "reading." Under our experimental conditions, however, where the observer is accustomed to take a specific task seriously and responsibly, we find that a fairly rich and varied supply of S-I and O-I is at hand. We shall not be surprised, moreover, when we increase the puzzling and enigmatical element in our typed objects, as in these samples in Series II, to find that both the number and the importance of the S-I and the O-I have increased. There the verbal formula to "read" does not carry the observer far into the task. Now the O-I take the forms "Does that make sense?", "That does not go," "What a puzzling thing!", "There is a real word: that is a beginning," "Now what is this word before it?" Frequently the occasion blocked the verbal instruction to "read" and substituted the O-I, "Is it a puzzle to be solved?" "How is this to be read?" and the like. At times the immediate and persistent blocking of the "reading" functions turned the verbal command into a haunting master-instruction of the self-type which meant "I *must* manage to read the thing." The chief form of the S-I came as goads, hints and encouragements. "There is a real word; now get the next," "Go on from left to right and see what you can find," "Try to group the letters," "That is not sense ('O Nan, do'); try another combination," "There is the end-word; now go backwards," "I must run through again," "Ah, I am getting the whole thing." Frequently the S-I was loaded with "experimenter's conscience," which charged the whole procedure with such emotive flavors as "I must hasten," "I am not doing well enough," "Stupid ass: why don't I get on!" And there was not wanting the "pious self" (every laboratory has at least one) which commented "I could have arranged matters better," "I will do my own way," "This is nonsense," "Why bother to analyse: I am a behaviorist."

A perusal of these typical S-I and O-I will make it apparent that the integrated and smooth "reading" complex was thoroughly broken up. What replaced it? Our protocols give us both the functions and their products. There is a good deal of what Kingsley (this volume, pp. 52–53) has called

"perceptive search." Instead of words bearing symbolic meanings the observer, driven by occasional and self instructions, sought out literal integrations which should "make sense." Presently came a partial comprehension, then a rejection, then more comprehension, then more "search," then more significant words, and finally the total meaning. On the way came, as in Series I, an occasional action (looking backward or forward under an S-I or O-I) and, more frequently, emotively-toned trains (vexation, remonstrance, self-condemnation, despair, reproach, flurry, etc.). Of true elaboration we found none. Once the troublesome words came out and went together, O discharging his task. Nothing was left to "think out"; or at least the O's did not instruct themselves to set thought-problems, as they frequently did in Series I.

Fractionation: Although simple our "reading tasks" brought out a fairly large amount of reportable stuff. We did not seek anything like the "complete introspections" recently criticized by Pratt[14] and long standing under suspicion. Our instructions did turn out, however, to be fairly varied and compendious. Still we cannot assume that the reports were exhaustive under our rubrics. There was, in particular, a tendency for the observers to translate their experiences into verbal associative trains and then to reconstruct, at the end, the individual instructions and to schematize the resulting functions. We therefore resorted to fractionation. A given sentence (of the kind previously used in Series I and II) was exposed for 1, 2, 4, or 6 seconds under the same conditions as before. New sentences were employed, so that the actual problem was fresh and untried. For the various fractionation times the sentences were given in unlike order and each of four Os (A, Ha, Ho and M) made 20 observations at each of the 4 fractionations; 320 observations in all. The numerical results follow (Table II):

[14] PRATT, C. C., *J. of Philos.*, 1924, xxi, 225–231.

TABLE II

Instr'n	*Exposure Times*	*1"*	*2"*	*4"*	*6"*	*Total Instr'ns*
Formal	Obs=A	20	20	20	20	80
	Ha	20	20	20	20	80
	Ho	20	20	20	20	80
	M	20	20	20	20	80
	Totals	80	80	80	80	320
Self	Obs=A	25	25	15	25	90
	Ha	15	17	18	22	72
	Ho	27	26	30	26	109
	M	9	13	21	21	64
	Totals	76	81	84	94	335
Occasional	Obs=A	1	4	16	11	32
	Ha	2	5	8	5	20
	Ho	4	11	6	11	32
	M	1	6	10	7	24
	Totals	8	26	40	34	108

It appears that, notwithstanding the brief exposures, about as many instructions of the " self " and " occasional " kinds resulted as in the longer, unfractionated trials. It seems probable, then, that the latter went far beyond the O's limit of report, thus losing a large part of the commentary. In the fractionated periods S–I are more frequent incentives to functional performance than the F–I given in the command (335 against 320). It also appears that O–I are relatively infrequent (108); and almost wholly wanting (8) with the 1-sec. exposure.

Intimations of individual tendency toward the form of instruction suggest that A and Ho were most inclined to substitute or supplement by inserting S-I and O-I. Together they contribute about 2/3 of these two classes. Psychologists of wide experience in analytical problems will have observed a tendency among untrained and instable observers to debate and to interpret in a variety of ways the fixed verbal commands which they receive. How little the mere words themselves actually determine function and output in the average " test," where means and methods are prescribed only in the vaguest and most ambiguous manner, may be conjectured. The variety and diversity in S-I and O-I may be taken from the following samples.

Obs.=A " What more should I do?" " Try hard to read it," " What does it really say?" " I will just look at it," " Why try really to understand," " No use; it goes too fast," " Do I believe that?" " I'll go back now to the earlier words," " Leave that word and go on," " Be quick," " Shall I try to answer the question?" " Does that make sense?"

Obs.=Ho "I must understand," "I will just perceive the letters," "I am not obliged to understand," "I can read this," "I am not satisfied," "I have done this one once," "Is this text sane?" "I no longer notice the formal instruction," "Nothing here but obscure perception of letters," "I shall hurry over the first part," "What does that word mean?" "I know this: I can be lazy," "To 'read' means to 'read words, not letters,'" "There *must* be something coherent here," "I wish that I had more time."

Obs.=Ha "I am trying to understand the whole," "I wish that I could have it again," "Those are just words, not sense," "He is playing a joke on me," "Are his instructions arranged in regular order?" "Don't bother to read beyond the first two words," "I wonder whether this one will be stupid?" "Well; I have conquered one," "I must get this one in the same way," "I won't attempt a solution," "I want to agree with that, but I cannot."

Obs.=M "I wonder if this is true?" "Only a jumble of letters," "I question its significance," "No need to include that," "There; I have understood," "There is the same sentence again," "Do not try too hard," "Perhaps you misunderstood," "Too slow," "Be careful," "Guard your attention."

Series III. Puzzling search. Throughout the first series the psychosomatic functions chiefly elicited were comprehension and elaboration, with more of the former than of the latter. In the second series, with shorter times and more hindrance offered to solution, the functional integration of "symbols," which we commonly call "reading," fell apart, throwing out elaboration and inducing fragments of the simpler functions; namely, perception, flashes of memory and emotion, and small impulsive actions. In order to call out a wider range of functional resources than in either of these series and then to see by what sort of instructional means these functions were brought into commission, we continued our study with more complicated and more "puzzling" situations. Of these we tried three varieties. Our results are too few for general inductions; but they paid for prospecting in this fertile field. The three varieties were (a) the match puzzle ("Arrange these five small sticks in such a way as to construct two closed figures, one of which shall be quadrilateral"), (b) the code problem ("Try to decipher the code"), and (c) the nonsense sentence which made sense only when turned over, held to the light, and read through the translucent paper ("Read the following message"). Most of the characters in this last puzzle were digits; but when seen through the page from the back the clear message appeared "Send ten troops on

the noon express." The "code" (b) consisted of 13 nonsense characters. The equivalent of one of these was given as "t" and of a second as "his." Seven observers were given (a) and (b), and six of them tried to solve (c)

(a) Here the initial formal instruction ("Arrange, etc.") first developed toward perception (the sight and tactual knowledge of the five sticks) and only slowly toward comprehension (*e.g.*, "so that is to be the construction"). Then followed various instructions.

S-I: "How have I done this kind before":—memory (Obs.=G)
O-I: "They are to be moved *so*":—action (Obs.=K)
S-I: "I must solve it":—comprehension (Obs.=W)
O-I: "What are the geometrical possibilities":—elaboration (Obs.=M)
S-I: "I'll begin with a square":—action (Obs.=A)
S-I: "I am failing":—emotion (Obs.=Ha)
S-I: "Just how am I to take his command":—comprehension (Obs.=Ho)

This sort of task, then, led to a wide variety of instructions and these instructions, in turn, to many functional operations. The fact that relatively more of the instructions were of the "Self" kind and fewer of the "Occasional" kind than in the previous series suggests that in such dilemmas the organism drives itself toward suitable functional means by first representing itself as an empirical and efficacious agent.

(b) The chief addition which came from the series of nonsense characters, to be deciphered by the aid of a fragmentary code, was the constant reference by the observer to the formal instruction and to the code. The organism was not equipped with the means for solution, and so it had again and again to seek in the formal task and in the characters given for "t" and "his" the clue to the meaning. Here a progressive comprehension of the characters came to be the end *immediately* sought, and the instructions thereto were many and various. On the whole the S–I form was relatively increased. In one series of 33 instructions (5 Os) 21 were S–I.

F-I: "Just what does the formal instruction mean?":—comprehension (Obs.=G)
F-I: "The formal instruction means 'These peculiar characters are to be read'":—comprehension (Obs.=K)
S-I: "I must refer again to the key":—comprehension (Obs.=K)
O-I: "What does '+' mean?":—comprehension (Obs.=W)
S-I: "I must learn these four letters":—memory (Obs.=Ha)
O-I: "Here are symbols without a code; how can I do them?":—comprehension (Obs.=Ho)

(c) The formal instruction was most ineffective of all in the trick message, which was read only by holding the paper toward the light in a reversed position. Simply to " Read the following message " was quite impossible. The dubious half-understanding of the command was followed by various O–I, hints from the nonsense sentence itself (" Has this a key? " " Is it similar to the preceding message? " " Possibly this period is meant to be a decimal point," etc.). As these O–I failed to bring comprehension, the observer had recourse to S–I (" I must do something," " How long am I to try? " " May I ask for help? " and the like). No one of these instructions leading to the desired comprehension, a real predicament developed and the trial ended with an emotion.

Summary: To study the forms of instruction and to observe the dependence of functional operation upon instruction we halted the prompt and high-organized " reading " function (Perceptive-apprehension-action-comprehension), which flows directly in the literate observer from a single formal instruction (" Read ") or a single occasional instruction (" Here is the page "). A large number of S–I and O–I then appeared, leading by more or less devious ways toward the task first set.[15] Out of these intercallated instructions came representatives of all the psychosomatic functions; namely, perception, memory, imagination, action, emotion, comprehension and thinking or genuine elaboration. While the results are by no means definitive, they illustrate a suitable method of approach to the experimental study of instruction. They display also the very wide variety of means for throwing into commission the resources of the organism.

Aside from the factual information about the amount and variety of instruction present on a given occasion and the relation of instruction to subsequent performances of the organism, the experiments bear upon several important psychological matters. We take them in order.

1. The correlation of " stimulus " and " response " by sound scientific methods is notoriously unsatisfactory. This one-to-one

[15] For our present purposes we may neglect those items in the commentary which indicate a ***procedure in course*** instead of introducing a new operation.

connection is as methodically bad as the older *Konstanz-Annahme,* which set the individual " sensation " into a state of unequivocal dependence upon the individual " stimulus." Against this formal correlation the *Gestalt-Psychologen* and others have, as every one knows, entered a vigorous protest. In the case of stimulus-response, one chief defect lies in the interposition of many variable and uncontrolled factors; factors which are best brought to light in instruction. On the physiological side is interposed, of course, the whole complex activity of the central nervous system. By a specious simplification the behaviorist ignores the " report " and reduces the intermediate bodily factors to a misconceived " reflex "; a simplification which appears more and more inadequate the closer we come to a knowledge of experience and the farther we recede from the naïve conception of " faggots " of neurones and " faggots " of sensations. The correspondence of two distant end-terms *a* and *z* is not likely to be made out either by a neglect of the alphabetical filling or by a theory of undisturbed propagation from end to end.[16]

2. The inference from output, as in the tests of performance, to " mental functions " is bound to be unwarranted so long as the " mental functions " are no more than hereditary names (" reading functions," " memory," " logical associations," and the like) set up to represent actual powers of " mind." It is only by the most carefully trained observation that we are likely to discover what actually precedes the turning out of a " test " result. Our study makes it clear that these antecedents are neither simple " mental functions " nor are they constant and invariable activities

[16] " For a long time psychologists were inclined to look upon 'mind' as a kind of hopper into which materials were poured by feeding in 'stimuli'; that is to say, by providing that the body should be carefully restricted to certain influences from without. Many behaviorists still take this older view, attempting to make a one-to-one correlation between 'stimulus' and 'response'; and the rough and uncritical procedure of the 'mental test' overlooks the fact that the means, as well as the amount, of accomplishment depends upon the exact way in which the organism approaches a given task. But the descriptive psychologist has learned his lesson. He is just as careful to control disposition and task as he is to standardize the immediate experimental situation. We may profit, therefore, by his discovery that the organism is not an idle machine which can be set into operation by pulling a lever; but that it always tends some-whither and always performs as it is functionally inclined."—*Field of Psychology,* 386–387.

of the organism. The human organism is extremely facile and plastic. Except in discharging its most highly habituated functions it pursues a variable and devious course to a given end. It follows that like quantitative scores neither prove that two organisms have operated in the selfsame manner nor that the operations have been introduced by the selfsame instructions and the selfsame bodily predispositions. Test results may, at times, be grossly prognostic; they are seldom indicative of the psychological conditions which brought them forth.

3. Our small prospecting study suggests that individual differences in instruction and in the means of performance have much greater *psychological* significance than have quantitative differences of output. By observing the way-in-which the organism goes about a given performance we seem to approach the very essence of individuation. Habituation appears, so long as it takes a fixed course, as a leveler. It tends to mask individual uniqueness. But when the psychosomatic functions are halted and checked we offer an occasion for the organism to discover its own means and methods. There structural inclinations, the accumulations of experience, the effects of specific education, the momentary state of the entire organism, and so on, are likely to exert their combined influence upon the actual performance which issues from the dilemma. The observations made under our method require, it is true, training of the order and precision used by the competent microscopist and mineralogist; a degree of training indeed which we did not wholly compass in our preliminary experiments. To know exactly "what is there" and to be able to report a matter-of-fact in the best possible terms and under the only appropriate categories is one of the two major results of scientific training in any subject.

4. Our wealth of "informal" instructions emphasizes the warning previously made against the assumption in many descriptive and psychophysical experiments that a good intent on the part of the observer is a sufficient guaranty that his reports are made under the steady influence of the "formal" command.

In fine, we have sought, by means of a standardized report

upon " instruction," to discover the exact conditions under which the total or psychological organism approaches its functional performance when confronted by a task. We have found these instructions to be more numerous and more varied than experimental procedure commonly assumes. They are conditions of functional output which are almost wholly neglected by the " test methods " and by the behavioristic correlation of stimulus with response. It appears that they should also be taken with greater seriousness by the descriptive psychologist, who is inclined to rely too confidently upon the formal command as the essential determinant of a psychological operation.

SEARCH; A FUNCTION INTERMEDIATE BETWEEN PERCEPTION AND THINKING *

By Howard L. Kingsley

Table of Contents

I. INTRODUCTION

Perception and thinking have long been outstanding topics in the study of the human mind. Standing as fundamental categories in man's reflections upon himself and his experiences, they

* The research was carried out under the direction of Professor Bentley, Director of the Laboratory.

have commonly been set apart in sharp contrast. Perception, on the one hand, has been regarded as a form of passive receptivity having a limited sphere of operation and dealing with the sensuous and earthy; while thinking, on the other, has been magnified and extolled as man's worthiest endowment, as an active, creative accomplishment capable of transcending the senses and attaining the spiritual and sublime. This traditional distinction, growing out of the ancient philosophies,[1] and flourishing under a now discarded doctrine of soul and faculties, must—if it is to survive in modern psychology—secure the sanction of experimental evidence.

The introduction of experimental psychology during the last century has changed the whole complexion of the discipline; but naturally enough the old ways have furnished the avenues along which experimentation has taken its course. Taken in a general way, the experimental work already done on thinking and perception has tended to break down the sharply contrasting distinction which time has fixed between them.[2] It has failed to substantiate those novel and unique "mental elements" of thinking which a number of investigators sought;[3] and, on the positive side, it has indicated certain points of similarity between perception and thinking with regard to the mental factors employed.[4] We can at most point to a general difference in the relative amounts and frequencies of the sensational and sensimaginal

[1] Cf. Baldwin, J. M., *A History of Psychology,* 1913, 33, 41; Brett, G. S., *A History of Psychology, etc.,* 1912, i, 30, 134; ii, 75, 77, 201; Locke, J., *An Essay Concerning Human Understanding,* Phil. Works (Bell), 1892, i, 253; Bain, A., *The Senses and the Intellect,* 1868, 8, 321, 611–667; Sully, J., *The Human Mind,* 1892, i, 212, 388f.

[2] See Titchener, E. B., *Experimental Psychology of the Thought Processes,* 1909, for an account and discussion of the work of Orth, Watt, Ach, Messer and Bühler.

[3] Bühler, K., *Tatsachen und Probleme zu einer Psychologie der Denkvorgänge,* 1907; Stout, S. F., *Analytical Psychology,* 1896, i, 85f; Woodworth, R. S., Imageless Thought, *J. Phil., etc.,* 1906, iii, 703f; A Revision of Imageless Thought, *Psychol. Rev.,* 1915, xxii, 1–27; The Consciousness of Relation, in *Essays philosophical and psychological,* 1908, 489ff; Angell, J. R., Thought and Imagery, *Phil. Rev.,* 1897, vi, 648f, 534f; Imageless Thought, *Psychol. Rev.,* 1911, xviii, 295f; Dürr, E., Ueber die experimentelle Untersuchung der Denkvorgänge, *Zsch. f. Psychol.,* 1908, 11, pt. I, 313–340; Clarke, H. M., Conscious Attitudes, *Amer. J. Psychol.,* 1911, xxii, 214–249.

[4] Titchener, E. B., *A Textbook of Psychology,* 1911, 48, 364; *A Beginner's Psychology,* 1917, sec. 25; Rogers, A. S., An Analytical Study of Visual Perceptions, *Amer. J. Psychol.,* 1917, xxviii, 546, 573, 576; Bentley, M., *Ibid.,* 239.

qualities,[5] a difference again which does not bear out the sharp distinction found everywhere in the earlier history.

A survey of the history of this distinction reveals the fact that it has been based in part upon the alleged "passive receptivity" of mind in perceiving, as contrasted with the alleged "active and creative" powers displayed in thinking, and also in part upon the notion that, while perception is dependent upon the physical environment and the sense organs, thinking depends only upon the mind or "soul." "Activity" and "passivity" as used in this connection are philosophical terms of only historical value. But something should be said of the alleged difference in the conditions of perception and thinking. No one will deny that the stimulus and the receptor are determining factors of importance in the initiation and organization of the perception, but observation has demonstrated that thinking is no more free from physical conditions. The difference apparently lies in the relatively greater influence of the central nervous system in the case of thinking, for here, as a rule, the receptors play a smaller part and the performance is characterized by a central determination in the form of the problem or *Aufgabe*. But we cannot sharply set perception and thinking apart by saying that perception is determined by outward factors and thinking by central conditions, for we know that the perceptive functions also are in part centrally determined. The nervous system is changed by the processes that run their course within it and these modifications or residues are always co-determiners of subsequent operations that take place within the brain.[6] "Every perception is shaped and moulded by the operation of nerve forces which show themselves neither in sensations nor in image. The nervous system . . . meets its impressions half way, and throws them into certain customary forms."[7]

On the other hand, we know that the conditions of thinking are not all central. For perception itself with all of its "outward"

[5] This term is used for the simple "centrally initiated" quality in the "image," Cf. Bentley, M., *The Field of Psychology*, 1924, 47.
[6] Bentley, M., *Ibid.*, 117.
[7] Titchener, E. B., *A Beginner's Psychology*, 1917, 115.

physical conditions not infrequently constitutes an immediate and fundamental part of thinking.[8] We also know that the influence of the stimulus and the receptor is not limited to the perceptive experiences. It indirectly determines to a fairly large extent both the qualities and the types of integration found in the subsequent imaginal formations.[9] Thus even the ideational material employed in thinking is not free from the influence of the impress of the environment upon the organism. It appears, then, that perception and thinking have much in common in the way of conditions as well as with respect to their constituent mental qualities. The differences actually found in regard to these factors would appear, then, to be rather of degree than of kind.

We find, as a matter of observation, that the fundamental difference between thinking and perception lies in function. Both are functions, operations, performances of the total organism. We cannot, then, expect a search for distinctive "mental elements" or for bodily dependence to be any more successful than the trite "active" and "passive" powers have been. Since perception and thinking are both modes of operation we must view and compare them as such. When we do this we find that perception is the apprehending of objects as present or of events as happening. It is thus an *apprehending* performance of the organism, an operation through which the organism lays hold of objects or becomes cognizant of their immediate presence and activities. Thinking, on the other hand, is found to be an *elaborative* type of performance. It is a working-out toward something unique by the use of symbols. It is a problem-solving procedure. Here, as it appears, we have an empirical difference between perception and thinking which serves to set them apart for investigation.

Our problem, then, is to explore by a suitable method these two types of functional operation in order to discover their fundamental similarities and differences, and further to ascertain

[8] Cf. Woodworth, R. S., *Psychology; A Study of Mental Life,* 1921, 469; Pillsbury, W. B., *Essentials of Psychology,* 1918, 229; Dewey, J., *How We Think,* 1910, 68–71; Bentley, *Ibid.,* 339, 366.
[9] Bentley, *Ibid.,* 141–142.

whether there may be intermediate modes which may serve to fill in or to bridge over the gap which tradition has fixed between them. We know that thinking is usually characterized by a stage of persistent groping for something not at hand. It moves by indirection toward a goal which is hinted at in the " thought-task " or *Aufgabe*. Now it occurred to us that in certain forms of perceptive scrutiny there is a similar factor of puzzling search in which the organism's mode of operation is not mainly determined by stimulus and receptor, and where something not yet " present " is to be half-uncovered and half-wrought-out. It was the " puzzling perception " which seemed to us to offer a possible intermediate between our extreme terms.[10] We proceeded, therefore, to set our perceptive and our elaborative functions into commission under comparative conditions and then to add our puzzling, prying variety under circumstances which would provide a direct comparison in both directions.

[10] Results of earlier laboratory studies on puzzle-solving and on " the elaboration of perceptual meanings " point to puzzling perception as something which partakes, to some extent at least, of both the nature of perception and of thinking. Cf. Rogers, S. A., An Analytical Study of Visual Perceptions, *Amer. J. Psychol.*, 1917, xxviii, 545, 551, 552; Mather, J. E., & Kline, L. W., The Psychology of Solving Puzzle Problems, *Ped. Sem.*, 1922, xxix, 269–282. The writer has had access to the results of two preliminary studies made in this laboratory on the " puzzle-solution " by Gerold C. Wichmann and Elizabeth Jane Rutherford.

II. THE EXPERIMENTS

In order to set into commission the three types of functional operation which our study undertook to investigate and to compare we presented pictures to our observers under fixed and formal instructions. During the presentation of each picture O was asked a question that pertained in some way to the presentation. The questions were of three kinds. The first type required only the perception of some readily observable part or aspect of the picture; the second called upon the observer to search for an obscure object, while the third type was designed to invoke real, though simple, cases of thinking. They are designated respectively as the perception-question (P-Q), the search-question (S-Q) and the thought-question (T-Q). The questions were simple enough to be readily understood and they were put as uniformly as possible in order to avoid any hint as to the type of operation expected.

The pictures used included photographic reproductions of well known paintings and suitable cuts from newspapers and magazines. The prints were uniformly mounted on gray cardboard, 7 x 9 inches, and all printed words were removed from them. Three sets of pictures were prepared. The first set included six with which P-questions were asked, six with which S-questions were asked and six with which T-questions were asked. Set II as made up of 24 pictures, six each for perception and search and twelve for thought. Six of the twelve thought-pictures were withdrawn before the question was asked. These are called the Ti-pictures to distinguish them from the others. In Set III there were four P-pictures, nine S-pictures and eleven T-pictures.

The exposure apparatus consisted of a wooden frame, $2\frac{1}{3}$ ft. x $2\frac{3}{4}$ ft., within which was fitted a metal screen covered with black cardboard. In the center of this screen was an opening, $7\frac{1}{2}$ x 5 in., fitted with an exposure shutter. This apparatus was mounted upon a table before which the observer was seated with his eyes two feet from and on a level with the exposure shutter. Large lateral squares of black cardboard were mounted at the sides of the exposure apparatus to restrict and simplify the field of vision. The experimenter stood behind the apparatus and out of O's sight. The mounted pictures were placed in the holders behind the opening in the metal screen. After the instruction and the ready signal the shutter was withdrawn exposing the picture.

The following instructions were given (save in the Ti-group): "At 'ready' you are to fixate. At the signal 'now' a picture will be presented and at the same time you will be asked a question about it. When you are able to answer press the button and immediately report your total experience during the exposure-period. Your report should be analytical and it should include in temporal order;

(a) The qualities immediately following the removal of the shutter;

(b) The qualities set up by the question;

(c) The subsequent qualities belonging to the period after the question, *e.g.*, visual, kinaesthetic, verbal-kinaesthetic and organic qualities and all sens-imaginal qualities;

(d) The relative clearness of the several constituents of experience;
(e) A full statement of meanings (always in parenthesis)."

The order of presenting the pictures of each set was determined by chance. No time-limit was set for the exposure-period. When O pressed the button a small light appeared behind the screen. This light was E's signal to close the shutter. O then wrote his report. The usual length of an experimentation period was about fifty minutes, during which time six observations were usually made, in some cases more.

In those experiments where the picture was removed before the question was asked (the Ti-group of Set II) an attempt was made to provoke thinking that would involve sensimaginal qualities derived for the most part from the preceding perceptual experiences. For the observations of this group the first part of the instructions was modified to read as follows: "After a 'ready' signal you are to fixate. At the signal 'now' a picture will be presented. When you are satisfied that you have fully apprehended the picture, press the button. You will then be asked a question about the picture. When you are able to answer the question press the button again and report your total experience, regarding particularly the period following the question." The latter part of these instructions was the same as the latter part of those quoted above.

During the course of the investigation it appeared to be instructive to observe the effect produced by repeatedly presenting a given thought-picture with the same accompanying T-question or a given search-picture with the same accompanying S-question. For a situation which today presents a problem for thought or search may, if the problem is solved, tomorrow evoke some other function. Repetition was used, accordingly, at unannounced points in Set III. This repetitive series we have designated as Set IV. The same instructions were given with them as with the original presentations. Two, three or four repetitions of a presentation were made at intervals of 2–11 days.

The factual material was contributed by five observers, designated as Y, Su, G, L and Sa. Three are members of the departmental staff at the University of Illinois. The other two were advanced students of psychology, apt and well-trained for the purposes of our study. These observers made a total of 210 observations, of which 58 were on perception, 64 on search, and 88 on thinking.[1]

[1] The writer wishes to acknowledge his indebtedness to all who have aided in this study. He desires especially to express his appreciation of the aid and counsel of Professor Bentley, at whose suggestion the problem was undertaken and under whose helpful direction and criticism this paper has been prepared. Acknowledgment is due to Professor C. R. Griffith for assistance and advice in planning the experiments as well as for his aid as an observer; also to Dr. Alice Sullivan and Dr. P. T. Young for their generous services as observers. The other observers to whom acknowledgment is due are Miss M. Libman and Professor N. W. Sanford.

III. RESULTS

All the observations, whether of perception, search or thinking, fall into four fairly well defined temporal phases. We shall find it convenient and feasible to present and discuss the facts obtained under these divisions. The first fractionation phase (designated f–1) extends from the opening of the shutter to the beginning of the question (Q). The second fractionation phase (f–2) covers the comprehension of Q. The third phase (f–3) is that in which O seeks and finds the answer to Q. The fourth (f–4) includes the pressing of the key.

A. *First fractionation phase:* In this part of the exposure period (from the opening of the shutter to Q) there was with all observers a hasty inspection of the picture. F–1 was, of course, common to the P, S and T observations. The instructions, formal and occasional, set up a tendency which led to the perceptive apprehension of the picture upon exposure. The *specific* course of this apprehensive operation, however, was not determined or directed by the formal instructions. These instructions provided only for a general " looking," while the particular object or aspect of the picture apprehended was left to central determinations. The duration of f–1 was approximately one second.

Of the experiential qualities appearing in f–1, the visual (blacks, whites, grays and browns) were the clearest and most prominent. These were arranged in a variety of patterns and carried the meaning of the picture as a whole or of specific objects within it. There were rapid and numerous eye-movements accompanied by shifts in the visual patterns as O observed now one part and now another part of the picture. Illustrations of these facts are given in the following quotations from the reports:

> Now one set of visual qualities and now another became clear as my eyes moved. (Su II 23 P.)[1]
> Visual perception of now one, now another, object in the picture. (Su I 10 S.)
> Visual qualities (grays, whites, blacks) changing in clearness as one passes from object to object (meaning: dog, seashore, ship). (G I 2 P.)

[1] Each report is designated by reference to observer, set (roman numeral), place in the set (arabic numeral), and task (P, S or T).

Strain qualities were also prominent. They signified eye-movements and general bodily sets. As constituents of an obscure somaesthetic background they frequently helped to carry the meaning of the picture as a whole, empathic experiences and mental postures. Examples follow:

Kinaesthesis, clear about eyes (active visual exploration). (G II 12 S.)
Kinaesthesis making up bodily set; the bodily set seemed to carry the meaning of doubt, surprise, etc. (G II 10 T.)
General bodily set (picture as a whole). (G II 17 T.)
Followed immediately by kinaesthetic qualities, partly shivers, partly pressures in the chest (meaning: I was in the wolf's place and very lonely). (Su I 17 T.)
Bodily set composed of diffuse somaesthetic qualities. (G II 15 Ti.)

Verbal kinaesthesis (V. K.) was mentioned almost twice as often as the non-verbal forms. It served to carry comments upon the general character of the picture and the naming of objects noted. For example:

V.K. very scrappy, faint (naming objects in inner speech). (G II 12 S.)

Sensimaginal qualities were rarely reported. When observed they appeared to supplement the sensational qualities in the perceptions of this period. They served to enrich the meanings of the pictures, sometimes by way of empathic experiences and sometimes by way of memorial reference. For example:

Visual perception accompanied by cutaneous and kinaesthetic imagery of coolness and dull pressure (meaning: I was out in the snow and was shivering). (Su I 11 P.)
Visual processes, a few imaginal and faint V. K. (apprehension of home and other objects, renewal in memory of visit to Mt. Vernon, etc.). (G II 3 P.)

Affective qualities appeared as characteristic constituents of the empathic experiences and were often found in relaxation. To quote from the reports:

A feeling of pleasantness blended with relaxation. (Su I 10 S.)
A feeling of relief and agreeable relaxation. (Y II 9 S.)
Visual qualities accompanied by a feeling of pleasantness and a certain amount of relaxation (meaning that I had put myself into the situation). (Su II 4 P.)

Before passing to the second fractionation phase we must consider the relation of the initial unrestricted exploration of the picture (f–1) to the more specifically determined operations following Q. The object designated in the P-question was usually

observed at once and often it was seen in f–1. More frequently there occurred in f–1 the apprehension of a complex object or group of objects, some part or aspect of which was asked for in Q. In such cases the P-task was simplified and abbreviated by the operations of f–1.

An example is found in the following observation:
Picture, "Old Charter Oak" (an unsymmetrical tree).
Q: "Is the large tree one-sided?"
Report (R), (a) Visual and V. K. (wind-blown tree). Very little visual exploration. Other objects in total picture carried in obscure background visual processes. (b) Auditory, (Q).
Processes alternated in clearness with visual processes sustained by a general bodily set (assent). (G I 16 P.)

The S-question usually instructed O to find some obscure object in the picture and since the preliminary survey familiarized him with various features of the picture it frequently prepared the way for and facilitated the search.

Picture: "Norwegians Making Hay."
Q: "Where is the baby-carriage?"
Report: One figure after another became clear. At the end of Q auditory perception was clearer than visual. Fixation moved to the baby-carriage almost immediately. (Su II 12 B.)

Again, in f–1, certain features of the picture were sometimes noticed which furnished clues for answering T-question and so the way was prepared for an immediate solution of the T-task.

Picture: "A Stampede," Rosa Bonheur.
Q: "What is happening?"
Report: Visual perception of animals (and a hazy meaning of stampede). Then clear auditory perception of Q. Next, a period of waiting in which I tried to decide just what was happening. First one part of the picture and then another came clear. I observed the man, then the animals and wasn't satisfied to say it was a stampede. This lack of satisfaction was meaning and I cannot say what processes carried it. Then I said in V. K. imagery, "It's a stampede," and pressed the button. (Why I decided just that way I cannot say.) (Su II 10 T.)

Picture: "The Ferryman's Daughter," Adan.
Q: "How deep is the water?"
Report: (a) Visual—rapid shift in clearness. Obscure V. K. (meaning: total picture-woman-boat). Very clear kinaesthetic imagery (as if I were poling the boat). Faint visual imagery (myself poling in a fast stream). (b) Auditory (Q), clear. (c) Q seemed somehow to fuse with the kinaesthesis mentioned above; made it clearer; led to a sort of bodily state (by which I read myself into the picture.) I "felt" the water to be about 2 feet deep. Judgment came in V. K. terms. (G I 18 T.)

Summary of f–1. Just before Q was asked there was a brief period during which O was occupied with apprehending the picture. O was disposed by the

formal instructions and the occasion to observe the picture but not any particular part of it. Visual qualities were clearest and most prominent. There were rapid eye-movements accompanied by shifts in visual patterns as O observed now one part and now another. Strains were prominent. They signified eye-movements and general bodily sets. Frequently they fused with organic qualities in an obscure background where they helped to carry the meaning of the picture as a whole, empathic experiences and mental postures. Verbal kinaesthesis carried comments and the naming of objects. Imaginal processes were not often reported. When observed they served to supplement the perceptions of f–1. Affective qualities were related principally to empathic experiences. The hurried inspection of the pictures found in f–1, which was common to the three types of observation, familiarized O more or less with the picture and so facilitated P, S and T in f–3.

B. *Second fractionation phase; comprehension of the Q:* It is obvious that through the formal, occasioned and self instructions that preceded the exposure of the picture a fairly intricate pattern of functional tendencies was set up. In f-1 we found that a tendency to scan the picture was released by the exposure. But there was also a tendency which set O for comprehending the awaited question, for he would be unable to carry out the instructions if he failed to understand what was asked. The first tendency was usually in complete control when the acoustic stimuli which released the second tendency were presented. That there occurred more or less of a clash between these functional tendencies is shown by the comparative obscurity of the words during the first part of Q, by the continuation of visual exploration of the picture during Q, by the oscillations of attention between the words of Q and the objects in the picture, and by the fight of Q for recognition. This clash was most pronounced when the time allotted to f-1 was insufficient for complete apprehension and Q interrupted or cut into the apprehending operations. It seems that the more completely the picture was apprehended before Q was asked the more readily was made the shift from apprehension to comprehension.

It is evident that the meaning of the question was the most commanding aspect of this part of the observation. Frequently this was all that O was able to report. It was immediate, unitary and clear, while " processes " were either not observed at all or very inadequately noted. For example, Y reports: " The words mean immediately, I do not hear them as bare sounds. I am aware of the meaning primarily." The meaning of Q was a

topic and its comprehension was in most cases an abbreviated performance, since the questions were usually short and easily understood. The topic usually emerged from the convergence of various factors, prepotent among which were the words spoken by E. The individual words of Q were not as a rule perceived. Sometimes, however, there were, while Q was being asked, attentive shifts to objects in the picture that were mentioned in Q. Here, then, visual perceptive meanings appeared prior to the topic. But the fact that so few meanings of this kind appeared seems to indicate that in most cases the words of Q immediately released specific neural tendencies which gave rise to the topic. For example:

> As Q is comprehended attention shifts from object to object mentioned in Q until the whole Q is comprehended. (G III 8 T.)
> Only at the end of Q did it occupy the focus. Until then it was in the background without meaning. (Su II 17 T.)
> Q was not clear until end, when it suddenly took on meaning and clearness. (Su II 18 S.)

There was a barrenness of experiential factors in this comprehensive operation compared with the significance of the meaning induced. Since Q was presented orally, auditory qualities were of fundamental importance. There occurred a shift, usually gradual, from visual to auditory experience. Most of the observers reported verbal comments, for this phase, which appear to make no direct contribution to the understanding of the question. Strains were reported from the chest and other parts of the body meaning " hesitancy " and " waiting " for Q " to come to an end," and also from the eyes giving hints of eye-movements accompanying the visual shifts which sometimes occurred during the course of the question. Those of the former group were obviously irrelevant to the understanding of Q and diminished as the Os became more accustomed to the experiment.

Summary of f-2. The visual exploration of f–1 was interrupted by Q which O had been disposed, by the instructions, to understand. Comprehension now succeeds apprehension. The more fully O had perceived the picture before Q, the more readily was made the shift from apprehension to understanding. The meaning of Q came suddenly or gradually. The dominant and clearest processes were auditory. In many cases process and meaning were not differentiated. In general, the meaning was clear, unitary, dominant, carried by a meager phenomenal thread. Several factors converged to set up the topic. The scarcity of meanings prior to the appearance of the topic indicates an important contribution by topical tendencies in the central nervous system.

C. *Third fractionation phase:* Now we approach the critical part of this analysis. Here we should find a difference—if it exists—between perception, search and thinking. The question once comprehended sets up a specific task in addition to the general instructions ("Look at the picture") of f–1. The P-questions were designed to set up the perception of a specific object within the picture; the S-questions to call out a quest for some obscure object, and the T-questions to invoke thinking. We shall first take up the description of the performances which followed the P-questions. These are designated as the P-functions, and the tasks thus carried out as the P-tasks. So with the other two types of question, performance and task (S and T).

(a) *Description of the P-functions; perception in f–3:* The P-questions disposed O toward the perception of some particular object or part of the picture. It is to be noted, therefore, that the perceptions of f–3 were more highly and more specifically determined by the experimental procedure than those of f–1. The objects named in Q were perceived in the same manner as those perceived in f–1, but with the added meaning in each case that the perceived object was the specified object. As a rule, so soon as Q was comprehended there was an attentive shift in the visual pattern and a kind of focalizing or limiting of the visual field. A particular part of the picture, that which was called for in Q, was singled out. This object was known at once, for there was a kind of identification which satisfied the demand made upon the organism by Q. This was followed immediately by the verbal formulation of a "yes" which led on to, or was accompanied by, the pressing of the signal button. The performance is characterized by its brevity, its directness and the promptness with which it is carried out, due, no doubt, to the simplicity of the task and to the immediate availability of all the factors essential to the accomplishment.

Both mental and bodily factors were employed by the organism in this performance. The clearest and most significant qualities for all observers were visual. Examples:

> Visual processes (dog) became clearer than any other processes. (G I 2 P.)

Apprehension of white house carried by clear visual processes. (G I 5 P.)
(A large development of branches to the right was noted), *i.e.*, visual processes—clear. (Y I 16 P.)
Then visual perception was clear again. (Su I 11 P.)

Next to the visual qualities, verbal kinaesthesis (at times imaginal?) is mentioned most frequently in P-reports. This served to carry verbal meanings, generally a subvocal "yes," and occasionally an irrelevant comment.

Other forms of kinaesthesis played an important part in the P-functions. K was frequently referred to the eyes, meaning eye-movement, visual exploration or strain of fixation. In some cases it was referred to the throat, and again it was reported as general muscular tension. Among the meanings reported as carried by the more general forms of kinaesthesis were satisfaction, relaxation, uncertainty, certainty, annoyance and ease.

Organic qualities appear to have played a comparatively inconspicuous part. "Bodily sets" are occasionally reported. These are usually constituted of a fusion of kinaesthetic and organic phenomena.

Y reports a "bodily attitude" made up of a pattern of pressures in the trunk, shoulders, and arms which conveyed the meaning "leaning forward to see better." L in one case reports organic pressures referred to the throat and stomach contributing "aversion." Sa describes pressures referred to the stomach and related to a desire to be in the place pictured.

Imagery apparently played a very small part here. It is rarely mentioned in the reports.

G reports a memory after-image of Q, and Y speaks of the words of Q lingering in imagery. L and Sa report a variety of imagery in this connection. This imagery, however, was not a pertinent part of the P-function but due to associative tendencies released by the apprehension of the picture.

Affective processes were not often reported. Pleasantness was reported in connection with satisfaction and relaxation which accompanied the answer to some of the questions. It was likely to appear when one of the P-tasks, all of which were comparatively easy, followed a distressingly difficult S- or T-task.

Mental postures are reported occasionally by most observers. They are usually unanalyzed experiences in which a relatively complex meaning is carried. Among the mental postures reported,

the experiences most frequently mentioned are the following: uncertainty, an attitude meaning "yes" or "true," "doubt," a "problem," "satisfaction," "assurance," "curiosity." The most obvious relation of the attitudes to the function under consideration is found in the service which they render as indicators of the progress made or being made toward accomplishing the task at hand.

As to the bodily factors that contribute to the perceptive functions in f–3, it is evident that besides the ocular resources the particular goal set by Q, as well as the general task carried over from the instructions, point to the presence of neural dispositions or determining tendencies within the central nervous system. There are also minor bodily contributions such as eye-movements and a straining forward to improve vision.

Summary of f-3 of the P-function.—Here, then, we have an apprehensive function which is more highly and more specifically determined (by Q) than the general apprehension of f-1. Q set for the organism a goal, the perceptive apprehension of a particular object. As soon as Q was comprehended, there were eye-movements and a shift in the clearness of the visual pattern. This usually was followed by the immediate apprehension of the specified object. Then appeared the subvocal answer to Q. Visual qualities were the clearest and most significant of the processes involved. Kinaesthesis of the throat muscles contributed verbal meanings, which appeared in the answers, and in irrelevant verbal comments. Kinaesthesis was a prominent element. It signified eye-movements and eye-strain, and it figured in such experiences as feelings of ease, certainty, relaxation. Organic qualities, fused with kinaesthetic, constituted "bodily sets" which sometimes meant assent to Q and the following of instructions. In a few cases, the memory after-image helped to carry Q over into f-3. The eye and the central nervous apparatus were the paramount bodily factors. The latter appears to have furnished a tendency which carried the organism through from Q to an appropriate goal.

(b) *Description of S-performance; search in f–3:* In the S-tasks, O was instructed to discover some certain object in the picture which was not likely to be observed by a casual exploration. Here, then, was presented the possible occasion for a functional performance of a new type. The comprehension of Q initiated a problem, and with the problem the organism was thrown into commission for a certain kind of operation; that is to say, the search for the specified object. The search was terminated by the discovery of the object sought, whereupon the organism was released. The duration of this performance varied with the degree of difficulty,—approximately from one to thirty seconds.

In the resolution of this task, O visually scanned, scrutinized and explored the picture. Visual apprehension with rapid shifts from object to object was characteristic of the performance. The visual qualities of experience were, therefore, numerous and by contributing the meaning of objects pictured, including the object of the search, they played an essential part. But the exploring and scrutinizing of the picture involved also certain non-visual qualities. Of these, muscular strain was by far the most in evidence. Verbal kinaesthesis was mentioned by observers G and Y on an approximate average of twice for each report on search; by other observers, less frequently. All of the successful S-performances were terminated by a subvocal "there," or a similar verbal meaning, carried by V.K., which served to announce discovery. V.K. also carried relevant comments made by O which appear to have helped him on his way toward the goal, and sometimes it aided in carrying into the period of search the instructions by way of a repetition of all or a part of Q. There were cases in which the name of the object of the search was repeated several times. To quote from the reports:

> V.K. (There's one person), (Where are the other two?). V.K. (There!). (G I 8 S.)
> V.K. becoming more scrappy but clearer (Coil-spring, coil-spring, coil-spring.) (G III 14 S.)[2]

Non-verbal kinaesthesis was found to play a conspicuous part in search. It is mentioned in practically every S-report and in most of the cases mentioned it seems to be relevant to the operation. A large part of the strain qualities reported in this connection was referred to the head, particularly to the eyes and the region of the face about the eyes. The K. of the eye-muscles signified eye-movements and combined with strains from about the face to carry the meaning of visual search, visual exploration, or active effort to find. These qualities were frequently supplemented by more general bodily strains and diffuse pressure qualities to bear the meaning of visual search. The "bodily set," represented in experience by the diffuse strains and obscure

[2] A coil-spring was what O was looking for in this observation.

organic pressures, furnished a background for the search performance and sometimes bore the awareness of carrying out the instructions. The termination of the effort to find the object specified in Q was usually marked by some form of relaxation or shift in the bodily set. Moreover, the reports indicate that relaxation is an important factor in the identification of the object, the discovery of which is the goal of the search.

Much K. about the eyes, general K. pattern about face, general K. of bodily setting, or attitude (search). (G I 8 S.)
K. pressures in head and neck (movement of head to and fro in endeavor to see man on horseback). (I was aware of eye-movement kinaesthesis during period of search.) (Y I 1 S.)
General K. strain in forehead and throughout upper part of body (movements of head and trunk to and fro as in looking). Pressures in eye (awareness of eye movement). (Y I 3 S.)
At Q, muscular pressure, weak intensity, and not very clear, waiting and search for the temple. (Su I 3 S.)
Vague scraps of K. (bodily set, instruction-following). K. about eyes (visual search). (G II 9 S.)
K. about eyes—general bodily set (visual search). K. about eyes clearer, general bodily set clearer. (Search of perceptual kind, exploration). (G II 24 S.)
There was bodily K. during this period (which meant that I must find the figure). (Su II 24 S.)
When Q came, my eyes went to the man who was stooped over and there was a pause and a slight relaxation (meaning: "He is the beast of burden"). Then when Q was finished with "which the peasants use," the attitude became active again, made up of organic and K. sensations, of strain localized in chest and head. This gave way almost immediately when I saw the donkey in the background. (Su I 7 S.)
Finally a group of processes became clearer. V.K. (There). General release of high muscular tonicity. (G I 8 S.)
V.K. (There)—(five). General change of bodily set (active to passive disposition). (G III 24 S.)
This gave way to relaxation (which meant my answer was formulated). (Su I 8 S.)

It appears from the reports that imagery was used very little in these S-operations. Observers G and Y occasionally found that the memory after-image and auditory verbal imagery served to carry Q into the period of active search. In only one of 57 S-reports do we find evidence that the observer experienced during the period of search a visual image of the object sought. Even in this case the visual image was accompanied by V.K. which carried the name of the object and because the image was not a true picture of the object it seemed to hinder rather than to

facilitate its discovery. It is evident that the factor which guided the organism in the quest grew out of the instructions.[3]

Affective tones were likewise not an important feature of the ordinary S-performance.

> Then a feeling of dissatisfaction because I saw nothing more. This ended in relaxation accompanied by slight unpleasantness and I pressed the button while pressure in throat meant (I can't find it). (Su I 15 S.)
> Slight unpleasantness soon begins to cover whole organic state (unsuccessful search). (G I 15 S.)
> Vague K. about face and slight unpleasantness (vexation that dog can't be found easily). (G III 13 S.)
> When I found I could answer the Q, I felt an immediate relaxation (this and the knowledge of being able to answer bringing with them the meaning of pleasantness). L I 1 S.)

Mental postures[4] (*Bewusstseinslagen*) were indicated by such terms as "exasperation" (G), "doubt-search-puzzle" (G), "ability to answer Q" (L), "doubt" (Y, Sa), "hesitation" (Y, Su), "uncertainty" (Y, Su), "keep on, you'll find him" (Su), and "wonderment" (Su). Thus the postures served chiefly as indicators of O's progress toward the accomplishment of the task in hand. These postures are usually reported by commentary, but in a few cases we have a partial inspective description of them. The processes reported are usually kinaesthetic pressures and strains referred principally to the eyes, forehead, throat and chest.

> In the meantime certain strains about eyes and forehead became clearer (doubt-search-puzzle). (G III 6 S.)
> At this time there was a weakly developed attitude of hesitation. (Y II 7 S.)
> This wonderment was carried by dull pressure in throat, but it was not as distinct as if I had actually said the words. It was as if I had made a shortcut in the nervous system so that a "humm—m—m" in the throat could mean sufficiently without really saying the words. (Su II 7 S.)
> Then I observed a mental attitude (which meant "yes") but which was not verbal motor imagery. It may have been carried by a bit of pressure in the throat, a shortcut for the complete act of saying "yes." (Su II 12 S.)

[3] Pyle found something similar in his analysis of expectation. "It is not the image of the coming impression, but the initial perception that throws the body into the attitude which gives rise to the characteristic consciousness we call expectation. In a very few of the experiments with any of the observers was the image a prominent factor. Even when images were present, they were seldom images of the coming impression." Pyle, W. H., An Experimental Study of Expectation, *Amer. J. Psychol,* 1909, xx, 561.

[4] Cf. Bentley, *Ibid.*, 351.

> This something was an attitude (meaning "keep on, you'll find him") but this was not said in words. (Su II 18 S.)
> This whole latter phase of the experience seemed merely a vehicle for the meaning (ability to answer the question). (Li I 15 S.)

Upon the bodily factors our study did not report directly; but certain relevant facts are obvious in the results. It is clear that the eye-muscles were essential and fundamental to the whole performance. Contraction of the extrinsic and ciliary muscles contributed numerous and rapid shifts in accommodation and fixation. These served to bring new and varied stimulus patterns to the retina which in turn were largely responsible for those rapid shifts in visual pattern which were found to be characteristic of the S-performance. The head was often moved from side to side and frequently also the trunk. Sometimes there was a leaning forward to improve vision. These gross bodily movements, as well as the various forms and abundance of kinaesthesis, indicate a rather widespread coöperation on the part of the general musculature. The bodily set held O to the task in hand until the goal was reached in the apprehension of the specified object. It was characterized by strained muscles and movements of the eyes, head and trunk. The mental aspects of the set are kinaesthetic processes, often obscure, resulting from such muscular conditions and movements; but the set is essentially and fundamentally a bodily condition. The neural mechanisms underlying are, of course, not revealed by the reports of our observers. But the facts indicate a functional priming of the central nervous system established by the Aufgabe.[5] It apparently is what carried the organism through from the comprehension of Q to the goal of the search.

Summary of f–S of the S-performance. The observer was instructed by the S-questions to find some obscure object in the picture. The problem thus set initiated and sustained a search. The organism was primed for the apprehension of a particular object and was released from the disposition to search as soon as the object of the quest was discovered. Rapid and numerous shifts in visual patterns were characteristic of the operation. V.K. was prominent. It was employed in naming objects, in the verbal formulation of the answer to Q, in sub-vocal comments, and occasionally it supported the memory after-image in carrying the instructions. The search meaning was carried

[5] Cf. Bentley, *Ibid.*, 384–396; Ach, N., *Ueber die Willenstätigkeit und das Denken*, 1905; Kries, J., Ueber die Natur gewisser mit den psychischen Vorgängen verknüpfter Gehirnzustände, *Zsch. f. Psychol., etc.*, 1894, viii, 133.

by kinaesthetic processes referred to the face and eyes, together with the general bodily sets in which were fusions of general bodily kinaesthesis and organic pressures. Sometimes the bodily sets helped to carry instructions. There was practically no imagery except verbal and the auditory memorial after-image of Q. Affective processes were occasionally present and served to indicate success or failure. Mental postures were frequent. They indicated more specifically progress toward the goal of the search. The bodily factors employed were the eyes, ocular and throat muscles and the muscles of the trunk and limbs; also, of course, the central nervous system. The latter furnished the determining tendency which carried the organism through to the solution of the task. In fine, we have, following the comprehension of Q: (1) an exploration in quest of the object designated in Q involving rapid shifts in the visual pattern and muscular strain; (2) discovery of the specified object accompanied by a sub-vocal answer (usually "there") and relaxation.

(c) *Description of T-performance; thinking in f-3.* The performance which worked out the answers to the T-questions has certain unique and distinguishing characteristics. The P-task was accomplished with the apprehension of a certain part of the picture. The S-task involved search and discovery on the perceptive level. The T-task involved both apprehension and search, and in addition, a certain amount of elaboration. The answer is not apprehended; it is thought out. It issues in part from the perceptive materials of the picture and in part also from associative materials. Q presented an occasion for thinking and the comprehension of Q as a rule gave rise to a problem which in turn initiated a search for solution. But the goal of this search was not a pictured object. It was rather an idea or opinion that should satisfy the demands of Q. It is clear that this performance was not "mental" alone. It involved both mental and bodily factors; and the scarcity and barrenness of experiential factors at certain critical stages is in many observations well marked.

It is evident from the frequent mention of shifts in visual pattern, visual explorations and visual apprehension that visual qualities of experience occupied a prominent place in many of the operations of this group. Frequently a visual exploration of the picture revealed clues which suggested the solution to the thought-problem. For example, the picture "The Lone Wolf" was presented with Q: "What time of day is it?" When O examined the picture with care he found evidence (symbolical meanings) in the stars and in the lights shining from the windows of the cottage.

White dots (meaning stars) became clear. They were clear and then obscured alternately several times. Inner speech "evening," then inner speech "night." (Inner speech was V.K.) Star dots again clear. An attitude of assurance not analyzed (meaning: It is night because the stars are out—that is right answer). (Y I 17 T.)

Clear K. (visual search). Mass of jumbled background. V.K. and bodily set (Are there any stars? Are there lights in houses? When do wolves come out? Are there any persons near? Are there any cattle about?). Along with the processes went visual exploration with running judgments carried automatically in V.K. (Two stars! No cattle! Lights in houses! It's about nine o'clock in the evening!) Judgment followed automatically an apprehension of the above objects—carried this time almost wholly in V.K. terms. Just a machine-like unfolding of (a) What time is it? (b) What are the signs of a time of day? (c) Are they present? (d) Yes! (e) Judgment. (G I 17 T.)

In several cases, however, clues to the solution of the problem were not found in the picture but in associative material. Then visual qualities played a much less important part. With the picture "A near view of the Colosseum, Rome," Q, "Why don't they repair this building?" was asked. Report follows:

K. about eyes changing rapidly in intensity as if eyes were in movement (doubt, why don't they, etc.). V.K. automatically (It's a relic). (It wouldn't be worth it.) General bodily set (hesitation, doubt). Sudden resolution of doubt—no previous warning—appearance of V.K. (Why! Who would think of touching an old landmark like that?) (G III 2 T.)

There is a marked abundance in the T-functions of kinaesthesis of the throat muscles carrying verbal meanings. V.K. is mentioned in practically every T-report and several times in many of them. It served, as in P and in S, to carry the answers to Q, subvocal comments and instructions. But a large portion of the V.K. reported for T referred to words, questions and comments concerned with clues and evidence bearing directly upon the solution of the problem. In other words, V.K. made a unique and effective contribution to the thinking operations in the form of "symbolical meanings."

V.K. and other qualities. (What clues are there to loud singing, open mouth, bodily posture.) Faint hints of muscular settings and V.K. (What would I do if singing loudly?) (G I 4 T.)

V.K. (Child walking away from house. Child wouldn't be out during or right after rain-storm.) (G I 9 T.)

Automatically formed judgment in V.K. terms. (They are going to the corral.) (G II 1 T.)

Memory after-images and V.K. (instructions).—V.K. (What evidence is there?)—V.K. processes (so large a figure couldn't be made of a single piece).—Judgment automatically came in V.K. terms. (G I 6 T.)

V.K., unclear (question)—V.K., unclear (must be mild). Immediate judgment (Spring). (G III 7 T.)

Background V.K. (a running comment on things that might be used). (G III 8 T.)
First a verbal "There is a storm. The wave raises the boat."—Verbal "There are heavy fish on deck." (This also seemed to answer Q.) (Y II 14 T.)
Then immediately in verbal motor imagery came the answer (Because the fish is so heavy). (Su II 14 T.)
V.K. (are they rocks?) (They are too close to land for icebergs to be so upright and relatively uniform. They resemble the smaller projection in the shore and therefore must be rock.) (Sa I 13 T.)

Non-verbal kinaesthesis is reported frequently though not so often as the verbal kind. It has a less direct bearing upon the progress toward the solution than V.K. has. It is chiefly concerned with furnishing a substantial background as well as certain secondary aids to accomplishment. The forms of kinaesthesis and the meanings furnished are varied. Thus G made 23 T-reports in which non-verbal K. was mentioned 22 times. It was referred to the eyes and forehead 16 times. Usually the meaning carried was that of visual exploration. Eye-movement was specified once. General kinaesthesis or bodily sets were reported 7 times; relaxation was given once. Upon three occasions, it is stated, kinaesthesis carried the meaning "doubt" and "puzzle." The other observers made fewer T-observations and did not refer to kinaesthesis as often as G did, but it is clear that kinaesthesis was an important part of their experiences. Frequent mention is made by them of eye-movement, attentive and bodily strains, and relaxation.

Diffuse and obscure organic qualities were occasionally reported. They apparently made no specific contribution to the thinking operations but coöperated with general K. in furnishing the background and general setting. A few cases of imagery were mentioned, the most noteworthy of which are memorial after-images of Q, an occasional pattern of visual sensimages bearing object-meanings relevant to the solution, and a case where kinaesthetic sensimagery furnished an empathic experience which helped O to arrive at an answer to Q. Simple feeling qualities were rarely reported.

All Os reported those flashes of meaning known as mental postures or conscious attitudes. Obs. Y mentioned them more frequently than the other observers. They sometimes carried a

comparatively elaborate meaning which served as the solution of the T-problem or which was employed in working out this solution. More often, however, we find that type of mental posture which serves as an indicator or gauge of the progress being made toward the solution. The meanings appearing in these indicatory flashes are expressed chiefly in the following terms: uncertainty, certainty, doubt, hesitation, uneasiness, assurance and dissatisfaction.

> Incipient verbal starts (How should I know. I don't know. They look like ice. They must be rocks because it isn't cold.) (P.S.—All of these meanings were present one after another, each one in a nutshell. They were not explicitly expressed.) The only part of the attitude definitely observed was pressure, weak, in throat and slight tension on forehead. (Y I 13 T.)
>
> Then a conscious attitude—not analyzed. (Its meaning might be elaborated in the form: The sphinx is made of many stones placed together—not a single stone. This was all given at once, *i.e.*, this meaning.) (Y I 6 T.)
>
> Attitude (I don't know. How should I know), not analyzed. Not explicit. Another attitude grading over from the first (meaning: There is only one house in the picture. The child must belong to it). (Y I 9 T.)
>
> The background processes lent a certain amount of uncertainty, perhaps discomfort, in the judgment as passed. (G II 14 T.)
>
> K. about eyes changing rapidly in intensity as if eyes were in movement (meant doubt—Why don't they?)—General bodily set (hesitation—doubt). (G III 2 T.)
>
> An attitude of assurance, not analyzed (meaning, it is night because the stars are out. That is right answer). (Y I 17 T.)
>
> Conscious attitude (not analyzed but meaning might now be stated as follows: both reasons are o.k., the storm and the heavy fish). (Y II 14 T.)
>
> A feeling of uncertainty as to the lights in the house; certain of the time, however. (Sa I 17 T.)

On the bodily side, the eye, face and throat muscles were employed in the thinking operations which resolved the T-tasks. The cerebral operations we can only conjecture. But the determining influence of the instructions, including the question, points to the existence of central functional tendencies set up under instruction which served to carry the organism through to the goal of the performance. In some cases specific mention of the determining effect of the instructions was made by the Os.

> Verbal experience (Why is the man at the stream?). (This seemed to determine my search, or to limit or restrict the search. There was a fixation upon the man pattern.) Visual pattern clearly delimited—rest of visual structure obscurely present. (Y II 13 T.)
>
> Then at the end of Q, the meaning came (I must decide this question; —although this was not said in words) and I began moving my eyes about. (Su I 17 T.)

> There was a total lack of puzzle or doubt, somehow a kind of lazy feelingless set to explore until end of path was found (a set which meant: there is an end to which it leads or he would not have asked it). (G II 17 T.)

Thus O seemed to be impelled toward a goal prescribed by the instructions and he was not released from the urge to find the solution of his problem until a meaning emerged which resolved the problem. The appearance of the appropriate meaning was commonly accompanied by muscular relaxation and a feeling of satisfaction which marked O's release from the problem-solving set. This is made evident by such reports as

> Judgment ("That's it") in V.K., terms not accompanied by the relaxation usually appearing. (G III 8 T.)
> Auditory was replaced by verbal motor imagery (meaning: "How do I know"). Noticed no tension or relaxation in this experience. (Su I 9 T.)
> After an instant, I pressed the button, but there was no satisfaction as there is when I can answer Q correctly. I think the dissatisfaction was mostly muscular tension in the forehead. (Su II 17 T.)

The duration and complexity of the T-performances varied greatly. In some cases, the solution appeared immediately, while in others it arrived only after much effort and many intermediate meanings.

Summary of f-S of the T-function.—Finding answers to the T-questions involved both apprehension and search, and, in addition to these, elaboration. The goal of this functional performance was not the apprehension of a particular object in the picture but the possession of an idea or opinion that would satisfy the demands of Q and that of necessity transcended the picture. Visual materials were prominent. The question and occasion disposed O to scan the picture. There were shifts in the clearness of the visual patterns as the eyes moved over the picture. Out of these visual patterns came meanings which were often symbolical, furnishing clues, to the solution of the problem. V.K. played an important rôle. It carried subvocal comments, clues, evidence, relevant questions and the answer to Q. Sometimes it helped to carry the instructions. The most conspicuous kinaesthetic qualities were localized about the face, eyes, and forehead, connected with eye-movements, and visual exploration. There were miscellaneous strains and pressures. These were sometimes found in the mental postures and empathic experiences. Imagery, outside of the verbal type and memory after-image which carried Q, was not often present and rarely aided in the resolving of the task. Mental postures furnished symbolical meanings and solutions, and served as indicators of the progress being made. Symbolical meanings were characteristic of all the T-functions. They appeared in the form of clues furnished by perceived objects and in the meanings of words used in the operations. Organic sensations were rarely mentioned. On the side of body, it is obvious that the eyes, muscles and nervous system were chiefly concerned, the latter by way of a functional trend. In general, after the comprehension of Q, we have (1) visual exploration of the picture, perceptions, inner speech, word-imagery, mental postures, memory after-image of words of Q; (2) answer to Q (the solution) in V.K. terms, or carried by a mental posture.

(d) *Description of the Ti-function.*—In Set II each O was given six tasks wherein the picture was presented under the instruction to press the button when it had been carefully scrutinized. No time limit was placed upon the period for examining the picture. When O pressed the button E closed the shutter and then asked the question. In solving this thought-problem O could appeal to the picture, therefore, only by way of memorial reference.

The first phase (f–1) of the observations of this group was substantially the same as f–1 in P, S and T. If there was anything unique about f–1 of Ti, it was the fact that O, anticipating the question, increased his efforts to make the apprehension of the picture circumstantial and complete. For example, Obs. G reports a bodily set which meant " see everything you can for you will be asked a question." The question was comprehended in these observations in the same manner as in P, S and T; and it served the same purpose as in the T-tasks.

The mode of operation in the solving of the Ti-problems was the same as that for the T-tasks. The difference lies in the fact that the visual qualities used in T were peripherally aroused while those used in Ti were centrally aroused. Relevant visual (sensational) qualities, prominent in T, were wholly lacking in Ti. Kinaesthesis of eye-movement and eye-strain carried the meaning of visual attention and reference to the recently scrutinized picture. Closely integrated with these were visual sensimaginal qualities which supported the same meaning. So complete was this blending that at times the ocular strains actually served to carry the visual meanings.

> Auditory (" Why is the girl standing under the tree? ") (Process and meaning together). With " girl " a visual reference (*i.e.*, I seemed to be looking. I do not know whether I had any image content or whether it was wholly meaning). Visual reference is a pattern of strain. Then scrappy visual content, a patch of gray with lines (meaning, rain). (Y II 5 Ti.)
>
> Auditory (" Why has the man raised his rifle? "). Immediately a pattern of facial strain—a pattern of attention—visual attention. Awareness of general bodily attitude. Then verbal inner speech, " because he is being attacked by wolves." Then (pressing key), with the attentive attitude there was a reference, very fleeting and scrappy. (Y II 15 Ti.)
>
> Auditory (" Is the boat approaching or going away? "), meaning immediately apprehended. With " approaching " and before the rest of the

sentence, there was inner speech "yes" and a vague visual reference was not clear (I seemed to be looking at the picture. I doubt if there was any sensory content present other than eye-pressures given in pattern of looking.) (Y II 20 Ti.)

A period at the first of the Q when no meaning was clear. Then the meaning of Q became clear and I observed a period of waiting in which I seemed to be deciding the answer. There were a good many shortcuts in this period and all that was attentively clear was a movement of the eyes causing a sensation of dull pressure (this *meant* the picture again; it was as if the eye movement bore a visual meaning). (Su II 8 Ti.)

Memory after-images, verbal experiences, mental postures and symbolical meanings rendered the same services here as they did in the solving of the T-problems. The following reports depict the general character of the performance:

Q in memory after-images, becoming rapidly less clear, and in V.K. becoming clearer, then fading away into a kind of bodily set (effort-puzzle-doubt-hesitation). Rapid review in visual imagery of objects in the picture. First, formulation of meanings not in original inspection (stag hunt). (Are they dogs or wolves? Is there a hunter near?) Shifting clearness of visual imagery. V.K. (There were only two dogs, no hunter). In background there were imaginal processes which meant (stag bleeding) but these processes could not seem to halt forming of decision—"Yes," which came "automatically," but rested somehow on the clearness of visual imagery and upon the vagueness of other imagery (meaning: bleeding). Had this last meaning become clearer, decision would have been reversed. (G II 8 Ti.)

The following report illustrates the general character of the performance and, more specifically, the manner in which a symbolic meaning carried by visual imagery may be the key to the solution. The picture ("Two Farewells," by Boughton) shows a ship in the distance and on the shore in the foreground are two women, one waving a handkerchief.

Q: "Is the ship arriving or departing?" The commentary is as follows: Auditory (is the ship arriving or departing?) meaning immediately apprehended. "Attitude" of perplexity. Attentive strains in forehead and possibly a diffuse pattern of bodily pressures (meaning: "Here I sit looking"). Then an obscure visual image of the picture. One pattern (white handkerchief in hand of woman symbolizing "Good-bye"—no words), became relatively more clear than others. (This is the solution). (Y II 21 Ti.)

Summary of Ti.—The procedure was similar to that of T. The picture, though withdrawn before Q was asked, sustained a similar relation to the function. The following modifications in function appeared, as revealed by the inspections and commentaries. While relevant visual sensations were prominent in T, they were lacking in Ti, but a corresponding place was filled in the latter by visual sensimages. Kinaesthesis, verbal and non-verbal, and mental postures appeared frequently and performed the same services as in T. The bodily factors concerned were also about the same as in T; namely, eyes, muscles, and central nervous system, the latter supplying the functional trend and conditioning the incorporations of the sensimages found in the memorial reference to the picture.

D. *Description of Set IV; repetitions of S and T.*—This set embraced a short series in which a "search" or a "thought" pic-

ture was repeatedly presented under the same instructional question. The aim was to discover if possible how the organism is affected by the recurrence of problematical occasions of this kind. The experimental conditions were exactly duplicated in the repetitions.

The foreground of picture II 12 S (Norwegians making hay) was filled with relatively large human figures, a horse, load of hay, etc. In the background were mountains and a village. These were the obvious features for apprehension in an ordinary brief survey. In one corner was an obscure and distant female figure pushing an infant's carriage. The Q ran: "Where is the baby-carriage?" The original presentation had involved the usual steps of the typical S-performance; namely, (1) survey of picture, (2) comprehension of Q, (3) search for carriage, and (4) discovery. Upon the second exposure, O (G) asked in inner speech (V.K.), "Have I seen this picture before?" Before Q was half finished, V.K. appeared with the meaning, "Yes—there's a baby-carriage." Thereupon, came the immediate subvocal "There." At the third exposure, there was an immediate apprehension of the picture as a whole, a background set, and scrappy, unclear verbal comments, "Oh! yes! That's the one we were talking about! There's the carriage." The object for which Q asked was apprehended before the question was presented. In the fourth exposure, there were clear visual patterns but no exploration. The visual and V.K. qualities together meant "Oh, yes! baby-carriage picture! there's the carriage!" This took place as soon as the picture was exposed and before Q was asked. With the fifth exposure, there appeared the usual visual qualities, but this time they were reported as obscure. There was clear V.K. and a bodily set meaning "That old picture, baby-carriage, there."

The picture for G II 18 S, was of a glacier, with mountains in the background and an obscure figure of a man in the lower right-hand corner. The Q was: "Where is the man?" Following the original presentation, there was a typical S-operation. In f–1 occurred the usual apprehending and naming of objects. This was followed by the understanding of Q. Next appeared K. about eyes and head, meaning visual exploration. This ended abruptly when the man was discovered and then appeared the usual "There" in V.K. Upon the second exposure, as soon as Q was grasped, there was an immediate shifting of the eyes, followed by a clear visual pattern (man) and a "There he is" in V.K. "In a brief after-period, there was a slight change in the bodily set and slight V.K. scraps (Why of course, this is familiar: I should have seen the man right away! I should have remembered the question)." The next time this picture was presented, there was, prior to Q, besides the apprehension of the objects a vague general set (I've seen this before). As soon as Q came, the "eyes moved as if I knew where the man was, but full recognition of the picture did not come until after the eyes had moved and the man was apprehended. It was as if the appropriate eye-movements had been touched off and the man apprehended before recognition of the picture appeared." With the fourth exposure, the glacier was at once clearly apprehended, there appeared V.K. (glacier) which carried also the meaning of the rest of the picture. There was no visual exploration, but a movement of the eyes and more V.K. (There's the man!), all of which took place while Q was still under way.

These observations make it clear that if an occasion sets up a search-problem and initiates a successful quest, the same occasion may not institute the same kind of performance if it occurs subsequently in the same manner. The search under the influence of

repetition at first becomes shorter and more facile. With continued recurrences of the same situation the problem and the search drop out and the individual comes to apprehend immediately without search. By virtue of the resolution of a search-problem, the object once sought may become the outstanding feature of a situation. We then have a perception determined largely by the outcome of the previous search-performance.

The reports show results which correspond to those found in search. At first, O tried to recall his former solution instead of trying to work one out. He was thus frequently relieved, by recall, of the necessity of thinking. As the pictures were repeated a second and a third time, the meaning which resulted from the former thinking became attached to the meaning which issued from direct apprehension. This meaning was carried largely by background K. and organic processes. The picture used in II 13 T was of a man standing by his saddled horse beside a river. The Q was: "Why is the man at the stream?" At first O (G) elaborated. But upon the second presentation he recognized the picture before Q came; and after Q he tried to recall his former answer. At the third exposure and before Q, there appeared the meaning, carried by background organic and V.K. processes, "I've seen this before; I must have a reason for the man being at the stream." The answer apparently was the result, not of thinking, but of recall of his previous solution.

For f-3, G reported; "V.K. becoming clearer at end of Q. (Oh yes, I have a reason. I gave it once—What was it?) In the meantime, background processes, organic and K. carried meaning (to go across, of course) and these led to pressing button; but the clearer V.K. were still present and merged into an unpleasant organic setting all of which meant (I have another reason, What is it? What's the use?, the judgment is already made, that's funny—becoming obscure—there must be a reason, etc.).

The picture for G II 6 T (Avenue of Trees; Hobbema) presented a view of straight rows of trees on either side of a road. Q. "Were these trees planted by man?" The original presentation set up a real, though simple, bit of thinking. The report for this observation is a typical T-report. With the second presentation and as soon as Q was begun, "there was a sort of organic change which meant (This is familiar: trees in a row: planted by

man; that is the question he will ask). Report immediately followed Q. The whole experience was crowded into a single unit. The principal processes were visual (with little exploration), V.K., organic and K., the latter carrying most of the meanings." As the picture was exposed for the third time, there appeared before Q clear visual apprehension of the trees, which seemed to carry with it the meaning of the whole picture, and V.K. (meaning, planted by man). Thus a new meaning became attached to the apprehension of this picture due to the previous solution of the thought problem. While in f-1 of the original exposure O apprehended the various objects of the picture, he later apprehended the trees *as planted by man.*

These reports on the repetition of occasions which originally aroused thinking reveal that: (1) The repeated occasion may or may not set up a problem. (2) Where the problem is set up, it may be solved by recall of the previous solution and thus no real thinking occurs. Memorial search may take the place of thinking. (3) Where a problem does not appear, the meaning of the situation may be enhanced by an attachment to it of the solution–meaning which emerged from the previous thinking. In this experiment, the answer sometimes appeared before Q was asked. It seemed to be linked with a part of the occasion which of itself was not directly responsible for the problem of the original presentation. Here, then, as in search, we have extended or modified perceptions due, in this case, to the outcome of former thinking. Such perceptions might be termed " sophisticated," to signify that they yield meanings deriving from former operations.

IV. DISCUSSION AND INTERPRETATION OF RESULTS

A. *Comparison of the P- and the S-tasks.*—The first two fractionation phases were common to P, S and T. In our comparisons then we shall deal almost exclusively with the third phase, f-3; namely, that part of the total performance which came after Q was comprehended. The tasks set by the P-questions and the S-questions are not the same, and the performances initiated are not the same. In P there is merely an immediate shift in the visual pattern to that part of the picture specified in Q. O then directly apprehends the answer. It might be said that O "recognizes" that particular object as the answer, the solution of his task, meaning by recognition merely that, in addition to the object meaning, there is the meaning that "this is the thing I was to note," or "this is what the experimenter asked about, now I must press the button." The specified object is there before him. He has merely to remark it. In S something different is done. A simple shift in attentive clearness does not here suffice. O must discover the object demanded. This is our "search"; a procedure sustained by a functional urge or disposition until terminated by the appearance of the specified object or (it may be) by a conviction that the quest is hopeless.

Visual perceptions occur in S as well as in P. In both, there is "identification," *i.e.*, the meaning that a particular object is the one demanded. But P is characterized by brevity, simplicity and directness, while S is slower and more complex, being characterized by muscular tension, effort, a striving for a specified goal and a visual exploration of the picture. It is, then, this sustained seeking for a specified object, a striving for a particular apprehension, that functionally distinguishes S from P.

The commentaries indicate that in S there was a greater number of visual patterns, a greater abundance of the verbal type of meaning, and much more kinaesthesis, especially K. of eye-

movements. Organic and affective qualities appeared more often in S. No significant difference, however, is found with respect to imagery.

The stages involved in P and S may be summarized as follows:

P-task.—(1) General visual survey of picture, (2) comprehension of Q, (3) eye-movements, clearness shifts, (4) apprehension of specified object and subvocal answer, (5) finger-movement upon signal key.

S-task.—(1) General visual survey of picture, (2) comprehension of Q, (3) searching of picture for specified object, visual explorations, many clearness shifts, much tendinous strain and eye-movement, (4) discovery and identification of specified object, verbal "there" (usually), relaxation, (5) key. The first, second and last of these stages are respectively alike. In P, the stages 3, 4 and 5 were brief, compact, and often telescoped. The stages of S were, as a rule, not so condensed. In P, again, the "answer" (stage 4) was usually a verbal "yes"; while that for S was most often a verbal "there," a difference traceable to a difference in the task assigned.

B. *Comparison of the S- and the T-tasks.*—Both S and T are determined performances. In both, the question asked sets up a specific problem in the observer who has been previously disposed by appropriate instructions toward a solution. The functional tendency thus established initiates, sustains and advances the problem. In both S and T, the observer seeks something not at hand, a meaning which will satisfy the demands of the problem. Since there is, as a rule, no experiential reference to the task during the period of search, it appears that an important contribution is made to both S and T by the central nervous system in the form of neural sets or functional trends. But the operation that follows Q is not the same in the two cases. In S, there is a visual exploration in quest of some perceptible object specified by Q. The goal of S is, then, a specific perception. In T, while there is also search and a striving toward a certain meaning, the goal is not, as in S, reached by the apprehension of an object. Here the goal is a bit of knowledge, an opinion or an idea; and it is reached

through an elaborative process which is made possible by the use of symbolical meanings. These meanings in turn are obtained through a functional fusion of search, apprehension, and understanding. Furthermore, the search which follows the S-question is restricted to the picture. A successful resolution of this task depends in large measure upon the ability of the observer to sustain a shifting of visual patterns until a pattern appears which bears the required meaning, *i.e.*, the object to be found. Sustaining this shift of patterns involves much ocular movement, changes of bodily posture and a considerable amount of muscular pull and tendinous strain. On the other hand, the search involved in solving the T-problem is not restricted to the picture. Visual exploration occurs, to be sure, where there is a possibility of obtaining through perception hints toward a solution; but often the search is extended to an exploration of O's memorial and other associative resources. This latter form of search we may call "ideational" to distinguish it from the perceptive variety. The use in T of ideational search implies that cerebral states and processes play a relatively more important part.

S and T have, moreover, much in common on the side of visual processes and visual meaning. In the reports of both, there is frequent mention of "shifts in the clearness of visual qualities," "visual exploration," "clear visual processes," "recognition of pictured objects," "shifts of attention," and "shifts of visual patterns," though a comparison discloses the fact that these visual factors are mentioned less often (a ratio of 2 to 3) in T than in S. Again, while the visual terms furnished in S the "solution" meaning, in T they furnished only clues or symbolical meanings which the observer used by way of elaboration to hit upon the solution.[1]

It is found upon examining the reports on V.K. and verbal comments that these experiences are given about as often in one as in the other set of reports. In both they serve to

[1] Visual exploration did not occur in all of the T-tasks for, in some cases, Q suggested no direct clue in the picture. Here the search was necessarily turned toward the exploration of imaginal resources or, more often, toward a fuller use of verbal symbols.

define the answer. For S the answer very often comes in the word "there," while for T there is a greater variety of verbal expression. Besides furnishing the answer and occasionally representing the task in the third stage, words in S usually had to do with naming and identifying objects in the picture; while in T they had chiefly to do with statements of fact gleaned from the picture or from the observer's knowledge and serving as symptom, clue or hint. Non-verbal forms of kinaesthesis are reported less than half as many times for T as for S. It is significant for this study to note that the most pronounced difference between S and T with respect to kinaesthesis is to be found in the local ocular variety meaning "visual search," "visual exploration" or changes of fixated pattern (48 times in 45 S-reports and only 18 times in 47 T-reports).

The stages involved in the complete observations for the two tasks under consideration are as follows:

S-task.—(1) general perceptive survey, (2) comprehension of Q, (3) searching the picture for object specified, (4) discovery and identification of object, verbal answer, relaxation, (5) finger-key.

T-task.—(1) general perceptive survey, (2) comprehension of Q, (3) search for symbols, (4) appearance of solution in verbal terms, (5) pressing key.

C. *Comparison of T- and Ti-tasks.*—As observed above, these tasks were very much alike. Both were elaborative psychosomatic functions. In the case of Ti the picture was not before the O during the solving operation as it was in T. Consequently there was a difference in the visual material employed. For in T peripherally aroused visual qualities were used, while in Ti a corresponding place was filled by visual sensimages. General and local K., verbal experiences and mental postures were about the same in both tasks. New meanings, not present during the original survey of the picture, appeared with the sensimaginal qualities under the influence of Q in Ti, just as new meanings appeared with the sensational qualities in T under the same influence.

Our T- and Ti-experiments indicate that in solving a thought problem the organism will make use of sensational material where it is at hand. Examples given in other researches, as well as our own T-performances, furnish ample evidence that perception does contribute to thinking. But in thinking the organism is not limited to sensory material. It can swing free from the immediate physical surroundings, when these are insufficient, and carry on its operations by means of kinaesthetic, imaginal and other resources. This fact and the one demonstrated by the comparisons of P, S and T, namely, that the organism uses like experiential factors in different modes of operation, support our previous contention that analysis by way of inspection alone does not yield a complete description of the organic performances in which experience is involved.

D. *Concerning the reports themselves:* Our instructions called for an analytical account, phase by phase, in terms of quality and meaning. To increase the adequacy of the report the typed instructions were either presented to O before each observation or left with him for constant reference.[2]

All observational reports we divide into the two forms of *inspection* and *commentary*.[3] Inspection is wholly confined to attributive terms. It never goes beyond "experience" (taken in the strictest sense), and it always describes experience in terms of its ultimate phenomenological character. But neither the total course of the experiential event nor its incidents and phases is reportable in inspective terms. These appear in the commentary

[2] Jacobson was among the first to succeed with a methodical procedure involving clear-cut parallel reports of process and meaning. In his paper "On Meaning and Understanding" (*Amer. J. Psychol.*, 1911, xxii, 552–577), he makes the following statement: "We do not here enter upon the question—which indeed is a question rather for epistemology than for psychology—how it is possible to give two parallel accounts, in terms of process and in terms of meaning, of one and the same total experience. The possibility has been taken for granted by previous investigators (Bühler, Dürr, Von Aster), and we simply follow in their footsteps. It should, however, be said expressly that the shift of attitude, from process to meaning or conversely, presented—after preliminary training—no insuperable difficulty to the observer. If a process or a meaning stood alone in the report, the failure was due to an inadvertence. All the observers found that duplicate accounts were possible, that processes could be summed up in a meaning, and that meanings could be paralleled by processes" (p. 556).

[3] See Bentley, *Ibid.*, 1924, 42–46.

which is made up, as the name implies, of the observer's coherent and first-hand "comments" upon the course of events. It is obvious that most of the reports from most of the psychological laboratories are not "inspective" at all. They stand instead under the class of "comments." In such a functional inquiry as ours, more material comes from the commentary than from inspection. It is quite as legitimate, when used methodically and with trained observers; and—in our case at least—it is wholly indispensable to a description of the functional operations of the total organism.

There was, however, enough inspective observation to denote the type and character of the materials involved. The inspective reports on P and T agree in general with like results of previous studies of perception and thought. This agreement speaks well for the reliability of the inspective descriptions of "search." Since our study was primarily concerned with a functional comparison of search with perception and thinking, however, the information given by way of commentary has been of the greatest value to us.

The presence of self-instruction as a factor in the reports is evident. This influence must be always considered in comparing the observers. Except under the most specific instructions, an observer's failure to report a given process or a given functional item is not a proof that the process or the item was absent. Self-instruction particularly affected the observations and the reports of f–1 and F–2. As the investigation proceeded, some of the Os seemed to get a hint that f–3 was the crucial part of the observation and therefore tended involuntarily to concentrate on that phase to the neglect of the others. There were enough detailed reports of f–1, however, to give us a fairly adequate idea of the functional activity of that period. The comprehension of the question was not reported fully enough to enable us to give a satisfactory account of it. The observers took it as incidental and usually reported it briefly or merely referred to the "question" without giving either inspections or comments. A study of comprehension did not, however, fall within our primary aim.

There are hints also of self-instruction throughout the reports on f-3. For example, affections are not reported at all by Y, whose chief experimental work has been devoted to the feelings, but frequently by L who was at work

on a problem in feeling at the time she observed for this experiment. The less experienced observers reported irrelevant processes more than the others. G's reports were more inspective than Su's. These facts must be taken into account in making quantitative comparisons. It will be seen, however, that the type of self-instruction noted is, for a given O, common to our P-, S-, and T-performances, and for that reason does not seriously affect our functional comparisons.

E. *Search regarded as a psychological concept:* In the resolution of our S-task we have found a type of performance resembling both perception and thinking but still possessing characteristics of its own. It is an active, attentive seeking. It is a quest, an anticipatory exploration initiated and sustained by a problem; and its goal is the discovery of a specified object or a defined end.

A preparation or disposition to perceive a certain object is common to both our P- and our S-tasks. But the fact that our perceptive operations were directed by the question upon a particular object does not render them less fit for the purposes of our study; for it is an established fact that all perceptions are determined in part by the stimulus and the receptor and in part by central or cerebral conditions. To ask a question that leads an observer to perceive a particular object may merely be adding a prepotent condition for perceiving. This may or may not initiate a search. Whether search follows will depend in part upon the nature of the *Aufgabe* and in part upon the availability for perception of the object specified.

The organism in searching looks toward the future; it anticipates. This it may do also in preparation for perception, as when the object is distant or delayed. But in search there is more than a disposition or preparation for a specific perception. There is more than anticipation, more than waiting, more than expectation. In search the organism aggressively goes after the thing to be apprehended with a view to discovering it instead of merely taking it as it is "given." The organism explores and hunts in an endeavor to discover. Perception may be used in the search, but the perception in that case is only a phase of the more complex searching function. Some of the searching operations of our experiments were much more difficult than others. The simpler and less difficult forms most closely resembled the perceptive functions. But even in these cases there was aggressive explora-

tion in seeking out something not at hand under the direction and urge of a problem. It is this feature that distinguishes search from perception.

On the other hand, search is like thinking in that both are determined, problem-solving performances. Both are initiated and sustained by a problem and both move toward a goal. Both exceed perception in that they reach out for something not at hand. But just as search includes and goes beyond perception, so thinking includes and goes beyond search. The search of our S-tasks has as its goal the apprehension of a specified object. The search found in thinking is functionally the same kind of operation, though its goal, instead of being a specific apprehension, is the derivation of symbolical meanings. Sometimes this search for symbolical meanings went on at the perceptive level, *i.e.*, the observer sought by way of visual exploration for signs that would suggest a solution. But this seen " object " was not apprehended as itself; it was used as a symbol. At other times, the organism explored its memorial and other associative resources in the same way for symbolical " suggestions " pointing toward a solution. The thinking operation is not finally consummated, however, with the symbolical meanings obtained by way of search. It goes further. The organism makes use of these symbols in setting up something unique, something new. Where search " finds," thinking " elaborates." Search brings the suggestions or materials out of which the thought solution emerges as the final step in the elaborative operations.

While the simplest forms of search appear most like perception, it does not follow that the most complicated forms most nearly approach thinking. A " difficult " search may last for a comparatively long time and involve a great number of meanings; while, on the other hand, simpler thought problems are sometimes solved in a flash. Search seems more nearly to approach thinking in those forms that do not have a specified perception as their goal. It is evident from this investigation that search may involve the use of the organism's associative resources, standing for its past experiences and knowledge, as well as the exploration of its physical surroundings. We may, as a matter of convenience, designate that form of search found in the S-task of this investi-

gation, in which the organism explores visual objects and reaches its goal by way of perception, by the terms " perceptive " search,[4] in order to distinguish it from those forms of search in which the organism explores its associative materials and arrives at its goal by way of memory, imagination or comprehension.[5] Where the goal is the derivation of symbolical meanings we may qualify search by the term " elaborative." The difference here between " perceptive " and " elaborative " is not a difference in mode of operation; rather it is a difference in the materials employed and in the outcome.

It is probable that we should come nearest the boundary separating search and thinking were we to call for an evidence or sign in the picture that would suggest a possible thinking solution. Here the goal would not be rigidly fixed by the Aufgabe upon a *particular* object; for the symbolical meanings might be attached to any one of several. Add to this the greater freedom in a search for symbolical meanings among " ideas " and we approach still more closely to real elaboration. But even then there is still the final step which is to be found in thinking and not in search; the step from the symbols to that which they suggest or create; to the solution of the problem, the new conviction, belief or idea.

If search does not give us perfect continuity of psychological operations, it does point to a real intermediary that relates perception and thinking in a functional way.[6]

[4] We speak of our own S-performance. Of course, " perceptive " search is not limited to vision.

[5] In report Y II 10 T occurs the following statement which bears upon this point: "Auditory, 'What is happening?' (Then a period of mental search, a groping for words.)—(P.S. This was characterized by a long period of suspense and doubt. The meaning was clear from the first, and it developed as I continued looking. The search was a search for words.)" Clarke made a study of the mental factors involved in " seeking " for the name of a forgotten acquaintance. The reports made by her observers and her generalizations resemble our own. After quoting several reports, she concludes: " It is evident that the consciousness of seeking consists of strain in the head and eyes or internal organs, and a feeling of effort localized in the head, together with images of any kind which have any connection with the required fact and would be likely to recall it." CLARKE, H. M., Conscious Attitudes, *Amer. J. Psychol.*, 1911, xxii, 214–249 (quotation from p. 227).

[6] In a recent volume Spearman refers to the " profound gulf which has from the most ancient time been declared to lie between sensory perception and abstract thinking." He maintains that there are transitional forms of cognition which effect a gradual passage from perception to thinking. See Spearman, C., *The nature of " intelligence " and the principles of cognition*, 1923, 78–81.

V. SUMMARY AND CONCLUSIONS

Throughout the history of psychology perception and thought have commonly been regarded as sharply contrasted aspects of mind. In this study we have undertaken to test by experimental means the validity of this tradition. We have approached our problem with the observation that first of all perception and thinking are neither mental powers nor mental "processes" but operations or functions of the total organism and that they are not sharply distinguished from one another on the basis of mental items. A distinction based upon the organism's mode of operation has seemed more appropriate. We then attempted to discover whether there was an intermediate mode of operation or function. It was noted that in thought there is a groping for something not at hand, a moving toward a goal suggested in the Aufgabe or problem. It occurred to us that in certain forms of perceptive puzzle-solving there is a procedure that is perceptive in character but which like thinking moves toward a goal suggested in the problem and that is not mainly determined by stimulus and receptor. It was this "puzzling perception," characterized by a searching and exploring similar to that found in thinking, that seemed to us to offer a possible intermediate function. Our experiments have yielded, by way of inspection and commentary, parallel descriptions of perception, thinking and this puzzling, prying variety of operation which we have called "search," carried on under comparable conditions.

Since the search found in the resolving of our S-tasks involves an exploration of objects physically present and includes and terminates with perception, we designate it by the term "perceptive search." The search found in our T-operations, on the other hand, we term "elaborative" because its goal is the derivation of symbolical meanings which are to be used in elaboration.

When we repeated over-and-over occasions for thinking and search, we came upon a significant foreshortening. In the case of search, the problem dropped out and search lapsed into perception.

In the case of thinking, the problem at first reappeared, arousing recall; but with continued presentations this problem too disappeared and the solution came at once. Thus thinking lapsed under repetition into search, and search into perception.

The results of the investigation seem, then, to justify the following conclusions:

1. Search is, like perception and thinking, a function or operation of the total organism. Standing midway between these two it serves to connect and to relate them.

2. The mental items employed in search are of the same kind as those used in perception and in thinking. There is here, then, no novel or unique mental "element" or mental constellation.

3. Search may include perception, but it also includes something more. While in simple perception the organism is engaged in apprehending present objects or on-going events, in search it is actively and attentively endeavoring, by way of anticipatory exploration, to discover a specified object or a defined end. In "perceptive search" this discovery is made by way of perception.

4. Search may be included in thinking; but here it is only a part of the entire operation. Thinking "elaborates" by the use of its symbolical meanings which are derived through search. This form of search is the "elaborative" form.

5. The intermediate relationship of search to perception and to thinking is further attested by the fact that thinking may be resolved under repetition into search, and search likewise may be resolved into perception—very much as choice under repetition drops to the impulsive forms and impulse in turn to automatic action.

STUDIES IN PSYCHOMETRIC THEORY

By Elmer Culler

TABLE OF CONTENTS

I. THE FECHNERIAN TIME-SPACE ERRORS

We shall attempt to prove that, along with the traditional time-space errors (p and q), first noted and named by Fechner, a third form of constant error (which we shall call *m*) appears when the receptive mechanism (hand or eye) moves twice to the right in passing from first to second stimulus and twice to the left. Although wholly ignored in the classic treatments of Fechner (8), Müller (19), Titchener (24) and others, m has precisely the same logical position and theoretical import as p and q.

In the method of constant stimulus differences, any response is affected with three forms of constant error, when the stimuli differ in temporal and spatial relation to the observer: time (p), space (q), and movement (m). Given two stimuli, a standard, S,

a comparison (judicandus), J; four time-space combinations are then possible, as denoted by Müller's four Hauptfälle (19, 64):

(a) S–J, left (J to the left and second)
(b) J–S, left (J to the left and first)
(c) S–J, right (J to the right and second)
(d) J–S, right (J to the right and first)

If we represent the four limens, which eventuate from these four combinations, by L_a, L_b, and so on, the time-error (p) equals $(\frac{1}{4})$ $(L_a+L_c-L_b-L_d)$; the space-error (q) equals $(\frac{1}{4})$ $(L_c+L_d-L_a-L_b)$; and the movement-error (m) equals $(\frac{1}{4})$ $(L_b+L_c-L_a-L_d)$. In terms of lifted weights, the three may be defined as follows: suppose the two stimuli, S and J, are equal; if the first one, whether to right or left, seems heavier than the second, p is positive; if the one to the left, whether first or second, seems heavier than the one to the right, q is positive; if J seems heavier when the second weight is to the left of the first, m is positive. In other words, if the gross (uncorrected) limen is larger: when J follows S, p is positive; when J is to right of S, q is positive; when the second weight is to right of the first, m is positive. The error is negative, in each case, when the reverse is true.

In these three are comprehended *all* the modes of error which the time-space relations of S and J can possibly induce. The first two, dating from Fechner, are known to all students of psychophysics; the nature and purport of the third (m) will here be briefly examined. Whereas the classical studies and statements of psychophysical theory make much of p and q, m is wholly passed by; they not only fail to accept it as a necessary form of time-space error on a parity with the other two, but they apparently fail to suspect its existence at all. We may well tarry a moment therefore with a constant tendency which has been so amply neglected in the literature of this field.

If we derive a limen for each of the four time-space combinations, the following equations hold (for convenience, we make the errors positive in the first):

$$L_a = L + p + q + m$$
$$L_b = L - p + q - m$$
$$L_c = L + p - q - m$$
$$L_d = L - p - q + m$$

L is the mean limen and equals (¼) ($L_a+L_b+L_c+L_d$), while the values of p, q and m are found with the formulae just given above.

Compare these equations with Müller's treatment of the matter, which is adopted and, in part, paraphrased by Titchener (19, 63 ff; 24, 272 and 302). We have, he says, a "konstante Gesamtfehler" c ($=\pm p\pm q$) which, as the word implies, embraces and sums into one total every kind of error induced by the time-space relations of the stimuli; he thus expressly fails to admit the existence of m. He now sets up equations for getting the true limen, "vom konstanten Fehler befreit," which in our terminology become (19, 67):

$$\begin{aligned} L_a &= L + c_a = L + p_a + q_a \\ L_b &= L + c_b = L - p_b + q_b \\ L_c &= L - c_b = L + p_b - q_b \\ L_d &= L - c_a = L - p_a - q_a \end{aligned}$$

The p and q (and their sum c) in L_a and L_d, as in L_b and L_c, are given the same subscript (a or b) in order to signalize that he considers them equal in magnitude, differing only in sign; there is "gar kein Zweifel" that we are justified in assuming p and q to be "absolut gleich gross" in a and d, as also in b and c; their equality may be taken for granted "ohne Bedenken"; for the reason that these two pairs are completely opposed, differing in time- and space-order both. When however two combinations, like a and b, differ in one order (time or space) alone, then, we are warned, it is "nicht statthaft" to assume p and q equal in size; they may indeed be of equal magnitude, but this is by no means necessary (19, 65). For example, a and b agree in space- and differ in time-order. We cannot here assume equality of the two p, inasmuch as "der Vorgang, auf welchem der Zeitfehler beruht, kann bei zuerst kommender linker (linksarmiger) Hebung von anderem Betrage und Erfolge sein als bei zuerst kommender rechter (rechtsarmiger) Hebung" (p. 66) No reason is given why the two q should differ, but he repeatedly asserts that neither one may be assumed equal in partially opposed orders (19, 65, 66, 69).

From his equations Müller finds the true limen, freed of all

space-time errors, in two ways: $L=(\frac{1}{2})\ (L_a+L_d)=(\frac{1}{2})\ (L_b+L_c)$. The two values thus obtained should agree closely; if they fail to do so, the reason lies with "absolute impression" or other factors which do not concern us here. But the errors p and q, he is careful to say, are indeterminable; c_a and c_b can indeed be found, but neither is separable into its two components. We could do so only by assuming with Fechner that p and q have the same absolute size in all four cases, which Müller calls an "unerwiesene und sogar überhaupt nicht kontrollierbare Annahme" (19, 70). Of p and q therefore we can say only twofold: (1) that a given p or q (as p_a or q_c) of an upper limen (say, heavier than S) is equal to the corresponding p or q of the lower limen (lighter than S); (2) that, in completely opposed cases, $p_a=p_d$, $p_b=p_c$, and so with q; about the relation of p_a to p_b and of q_a to q_b, however, we can say nothing (p. 70). We thus know various interrelations of the several p and q; but their actual magnitude cannot be determined in a single case.

What are the facts? We may state them as follows:

(1) L is equal neither to $(\frac{1}{2})\ (L_a+L_d)$ nor to $(\frac{1}{2})\ (L_b+L_c)$; on the contrary, it is the mean of these two quantities, that is, of the four empirical limens.

(2) L_a+L_d does not equal L_b+L_c; on the contrary, $L_a+L_d=2L+2m$, while $L_b+L_c=2L-2m$.

(3) None of Müller's four "konstante Gesamtfehler" (c_a —— c_d) is equal to any of the others; on the contrary, they all represent differing combinations of m, p and q.

(4) All of the four p or q or m are, by definition, equal; their values can be readily found.

(5) His assumption (after Fechner) that p and q are equal for both upper and lower limens (say, heavier and lighter) is unwarranted; this can be no more taken for granted than equality in the limens themselves.

To all these errors, m is the key; they vanish as soon as it is introduced to our equations. Professor Müller himself provides the evidence for m and thus confutes his own position, as a little

analysis will show. He tells us that we dare not assume equality of p when the first weight lifted is now on the right, now on the left; inasmuch as the process, upon which the time error depends, may be of different amount and consequence in the two cases (19, 66). He later confirms and elaborates the above statement. "Ein Fall, wo der linke Reiz zuerst gegeben wird, ist anders geartet als ein Fall, wo der rechte Reiz an erster Stelle kommt" (p. 69); in illustration of the statement, he points out, "das sich die Aufmerksamkeit der Versuchsperson dem Reiz der einen Seite, z.B. dem rechten Reize, mehr zuzuwenden pflegt als dem Reize der anderen Seite. Ist also der Grad der Einprägung des ersten Reizes mit von Belang für den sich ergebenden Betrag des Zeitfehlers, so würde man gelegentlich zu wesentlich unrichtigen Resultaten gelangen, wenn man diesen Fehler beim Vorangehen des rechten Reizes für gleich gross ansetzen wollte wie beim Vorangehen des linken Reizes" (p. 69). This holds not merely for mental factors like attention, but for purely physiological ones like fatigue as well (p. 70). In face of these unequivocal statements, he ventures explicitly to equate the two quantities, $L_a + L_d$ and $L_b + L_c$, even though they differ in precisely the way he mentions: in a and d the *right* weight is lifted first, in b and c the *left* one (p. 67, equations 11 and 12).

No more need be said to prove that we have reached an impasse. Müller well appreciates in theory that the psychological situation alters if we take the right stimulus now first and now second; but when it comes to applying the principle, he is apparently misled by his own equations ($L_a + L_d = 2L$, $L_b + L_c = 2L$); since both quantities equal 2L, they must equal each other. This unfortunate conclusion vitiates, in large part, Müller's whole analysis of the time-space errors. The equations, of course, are inadequate; for the reason that he fails to recognize that four time-space combinations *necessarily* induce three distinct modes of error, no one of which has logical priority over the others. By the ordinary laws of probability, six combinations are possible with four things (the four limens $L_a - L_d$) taken two at a time: ab, ac, ad, bc, bd, cd; whence we have three pairs which are opposed in one

respect while agreeing in the other two: limens ab and cd differ in the spatial relation of J and S, but include the same time-orders and relative position of first to second weight; limens ac and bd differ in time-order, but agree in the spatial relation of J to S and of first to second weight; while limens ad and bc differ in the relative position of first and second weight, but agree in time-order and spatial relation of J to S. Is there any logical or practical reason for adopting the first two of these and rejecting the third? None whatever! There is no more justification for the equation, $L_a + L_d = L_b + L_c$, than for $L_a + L_b = L_c + L_d$, or for $L_a + L_c = L_b + L_d$; all three are equally wrong: in the first the two members differ by 4m, in the second by 4q, in the third by 4p. On first thought, it may indeed look natural to assume one time- and one space-error, as Fechner did; but however natural, it is assuredly incorrect.

With p and q alone, therefore, we cannot do justice to the complexity of the psychological situation; a third tendency, m, is here operative and must be given room in our equations. In theory, as we have seen, Müller seems to recognize an additional factor; but he persists in treating it as a component of p, with which it has no more and no less to do than with q. In consequence he is forced to conclude that the four p are not all equal in magnitude; they change as the left stimulus is now taken first and now second; whereas all he succeeds in proving is that p+m is not identical with p—m. He nowhere gives evidence that the several q are unequal in size; he always resorts to p for illustrative purposes and then generalizes the argument to include q. Having made his conclusion, he proceeds to criticise Fechner for assuming the four p (and q) equal in magnitude. In so doing he states the situation wrongly: the several p and q (as also m) are equal, not by assumption, but by definition. We have just four observation equations (L_a——L_d); each equation has four unknowns (L, p, q, m); each unknown therefore is *uniquely* determined. The value of p is simply the figure which emerges upon solving the four equations for p, just as L is fixed by the figure we get upon solving for L. Our only means of knowing that there are constant

tendencies like p or q or m is that the four observation equations are not identical. It is quite beside the mark to speak of "assuming" that p is equal in the four equations; p can have but one value if the equations hold at all.

In illustration of correct procedure, we may cite an example from the writer's study of thermal discrimination (7). The four warm limens (L_a —— L_d) at 16° are .040°, —.034°, .179°, .059°; whence the four unknowns are found:

$$\begin{aligned} L &= \ \ .061° \\ p &= \ \ .049° \\ q &= -.058° \\ m &= -.011° \end{aligned}$$

Müller's (and Fechner's) assumption, finally, that p and q are equal for both upper and lower (heavier and lighter) limens, is without foundation. We have two different sets of observation equations, four with the upper and four with the lower. Obviously, there is no reason why p or q should be identical in two independent sets of equations; the limens are notoriously unequal in the two.

To sum up: when we use four time-space combinations, four limens are found; no one of these, and no combination of them, is equal to any other; each is affected with a separate combination of three constant tendencies, all of which follow with necessity from the use of two time- and two space-orders; and each can be adequately presented only by an observation equation with four unknowns, all of which are uniquely determined.

Müller's unquestioned authority in this field has led to general acceptance of his conclusions; but one may hope that *all* of the constant errors will hereafter be given due consideration in psychophysical theory and experimentation.

II. ABSOLUTE IMPRESSION AND THE " TENDENCIES OF JUDGMENT "

In this article we desire to point out and rectify an error in Müller's treatment of the judgment-tendencies, to develop a more adequate theory in accounting for them, and to propose a simple and reliable means of preventing them.

The doctrine of absolute impression was introduced by Martin and Müller (17, 45) to explain the " generelle Urteilstendenz " and the " typische Urteilstendenz." Müller defines the former as follows: " Es besteht im allgemeinen eine Tendenz, bei gleicher wirksamer Differenz mehr richtige Fälle zu ergeben, wenn das Vergleichsgewicht das zuzweit gehobene Gewicht ist, als dann, wenn dasselbe an erster Stelle kommt " (19, 115). The phrase " equal effective difference " (gleiche wirksame Differenz) is the crucial matter here, and may be defined as follows: When a given time-order increases the number of correct judgments, we may, as Fechner suggested (8, 115), regard it as enlarging the objective stimulus difference (D) by a given amount, p; the space error likewise equals ±q; so that the " effective difference " equals (±D±p±q) as opposed to the mere stimulus difference ±D. Consider the four time-space combinations, both when comparison (J) is larger than standard (S) stimulus, so that D is positive, and when J<S (— D):

	I	II
a.	+ D + p + q	— D + p + q
b.	+ D — p + q	— D — p + q
c.	+ D + p — q	— D + p — q
d.	+ D — p — q	— D — p — q

We find that each of the four in I is a direct negative of one in II; thus, (I, a)= —(II, d). In I, a, J exceeds S by (D+p+q), in II, d, S exceeds J by (D+p+q); we thus have the same " effective difference " and should get the same percentage of right judgments (a slight difference from the fact that the first stimulus in II, d is smaller, by the amount D, than in I, a need not concern us here). On the contrary, Müller found, in the case of all his

observers and with a variety of experimental procedures, that these pairs by no means give equal proportions of correct responses: time-order S–J on the whole yields more rights than does J–S.

A "typical tendency of judgment" appears when the sum of the right responses for the four combinations of I (+D) is not equal to that of II (— D). When the latter (— D) shows more rights the "type" is positive; when fewer, the type is negative.

Müller proposes to explain these facts with his doctrine of absolute impression, which may be summarized in the following propositions (19, 117):

1. A given judgment (heavier, lighter) may issue from an actual comparison of S and J, but may also eventuate merely because one of them seems "absolutely" heavy or light without regard to the other.

2. This absolute impression appears oftener with J than with S.

3. It appears oftener when D (= J — S) is large than when D is small.

4. An absolute impression from the second stimulus affects the response more than does one from the first; in the latter case it acts upon the judgment "nur mit Hilfe der Erinnerung" and is "zuweilen schon ganz vergessen" by the time judgment is reached.

5. Vigorous lifters (kräftige Heber) get an absolute impression of lightness more easily than of heaviness; with less vigorous subjects the reverse is true.

Three questions now engage us:

Is Müller's factual presentation correct?

Is his theory adequate?

How can the tendencies be prevented?

(1) What shall we say of Müller's analysis? Mainly that none of the four equations (I, a = — II, d and so on) is correct. We have shown in the preceding article (I) the existence and operation of a third time-space factor (m) which is on precisely the same plane as p and q; to pass it by as nonexistent or

negligible is quite indefensible. The true combinations for (+D) and (— D) are as follows:

	I	II
a.	D + p + q + m	— D + p + q + m
b.	D — p + q — m	— D — p + q — m
c.	D + p — q — m	— D + p — q — m
d.	D — p — q + m	— D — p — q + m

If we now try the equation, I, a = — II, d, we have: (D+p+q+m) = —(—D—p—q+m), which of course does not hold; the two members differ by 2m. This is true of the four pairs which Müller equates; indeed we are no more warranted in setting I, a = — II, d, which differ by 2m, than in putting I, a = — II, c, which differ by 2p, or I, a = — II, b, which differ by 2q. Whatever unlikeness in the proportion of right judgments is found between I, a and II, d must therefore be ascribed in part to the disparity of 2m in their effective differences; even were no judgment tendency operative, the two would be equal only when m is zero.

The factor m can be removed by appropriate combination: I, a+I, c = 2(D+p) = —(II, d+II, b). We are now ready to apply Müller's argument: the effective differences being really equal in these two equations, we may expect the same percentage of rights from each member, unless some other influence is helping determine the response. Referring to Müller's table (19, 114; Tabelle 5 and 6), we find that the proportion of rights in I, a+ I, c (time-order S–J) definitely exceeds that in II, d+II, b (time-order J–S). Our analysis therefore agrees with Müller's in constating the tendency to give more rights in the example just cited when J follows S than otherwise; the factor m is here relatively small; but our treatment aims to be adequate in theory, which can hardly be said for Müller's.

(2) What of Müller's general doctrine? It appears defective in no less than three ways: (a) the concept " absolute impression " is but loosely and casually defined; (b) it is credited with an import far beyond its demonstrated significance; (c) the doctrine explains nothing: it gives a name to a group of facts without trying to expose their underlying connection with related phe-

nomena of judgment. Let us consider these points in order. (a) To our knowledge Müller nowhere says just what "absolute impression" (a.i.) means. He refers to the common experience that a book or child may seem heavy or light without being compared with another specific book or child; so it is with lifted weights. But suppose, in a given series, we are comparing weights in the vicinity of 100 gm; 200 gm would then, without any question, seem "absolutely" heavy; but would 200 gm give the same a.i. had we been comparing weights of 300 gm just before? Obviously not. It is clear that the term "absolute" must be taken in a kind of pickwickian sense; the impression from 200 gm is wholly relative, not perhaps to a single other impression, but to a group or series of repeated impressions at least. Suppose weight A is compared with B, the two impressions combining into a single gestalt; certain criteria (intraorganic responses) appear which are characteristic of this mode of configuration and constitute what is called the perception of difference; whatever this means in neural terms (a flow of ions or difference in potential between the two members of the gestalt; *cf.* 16), the process eventuates in the response "heavy" or "light": this is what we mean by direct comparison. But sometimes the new impression from A forms a configuration with a *group* of residual impressions left over from antecedent stimulations and now more or less revived: this is what happens in "absolute impression." With a.i. indeed the criteria are often more stable and well defined, and the perception of difference accordingly less obscured with irrelevant factors, than with direct comparison. In any event the distinction between a.i. and one-to-one comparison is one of convenience and not of theory. The relative character of a.i. appears further in the fact that, other things equal, recent impressions have disproportionate influence upon a judgment; an "absolute" impression is obviously affected more by the weights we handled an hour ago than by those of yesterday.

Absolute impression therefore is relative to both the ***number*** and ***recency*** of antecedent impressions. Properly understood, the

term denotes a fact; but the name and usage Müller gave it fail to reveal its true character and its relation to direct comparison.

The intimate relation of these two serves to explain the difficulty of deciding when or to what degree a response is due to a.i. and when to actual comparison. We have no means of knowing save by O's verbal report ("introspection"); but Müller to our knowledge gives no results at all where O is required to announce for every case whether a.i. determined the judgment or not; while without this explicit and detailed evidence, Müller's explanation of either judgment-tendency is barren of any real proof. After protracted experience in comparing temperatures and some with lifted weights, we find *e.g.* that the first weight of a pair may seem markedly heavy, to be sure, and thus arouse a kind of expectation that the second is likely to be lighter; but regularly, and we believe invariably, the first seems "heavy" *in relation to* the pair or pairs we have been handling just before. If these preceding weights were, say, twice as heavy as they now are, does anyone believe the first member of the next pair would still seem "absolutely" heavy? When a first weight is unusually heavy we have asked again and again, why? to which we could only reply, because the last pair or two were much lighter. We have also tried reporting for each trial whether a.i. determines the response in case it appears with the second stimulus; and have never been able to find that the "absolute" heaviness or lightness of the second was independent of the first. It is almost sure to be affected of course by other antecedent impressions, as we noted above; the two factors (comparison with the stimulus just preceding and with others more remote) seem to combine in varying degree from case to case; but the former in our experience is invariably present.[1]

We are in short convinced that the concept a.i., as defined by Müller, is too elusive and casual to be usable for quantitative purposes.

[1] In a long series O may become weary and inattentive and thus begin to reply more or less at random; so that the first stimulus (and for that matter the second) may then have little influence upon his response. But in our discussion, of course, we presume that O is doing what he is instructed to do and what his responses, if they mean anything, presuppose he is doing.

(b) Does Müller's doctrine have the significance he ascribes to it? Consider the five propositions in which he applies it (p. 64). A.i. being defined as above, the first three seem acceptable, even obvious; but they have nothing to do with the judgment-tendencies which the doctrine is designed to explain. The fourth purports to account for the general Urteilstendenz but seems inadequate, not to say incorrect, for at least two reasons: (1) we have just seen that, in our experience, a.i. is never pure when it appears with the second stimulus; the first member of the pair always helps to determine and modify the total impression; (2) Washburn (27, 220) reports, when S follows J, a tendency to estimate the first stimulus absolutely; O does not wait for the second (S) but tends to report at once that J is "pretty small" or "unusually large." Müller admits that the effect of a.i. in this case would "sich der Selbstbeobachtung der Versuchsperson leichter aufdrängen" than when J follows S (19, 123). This fact well agrees with our conclusion that a.i. affects the response more when J precedes S than otherwise, but is hardly consonant with Müller's proposition 4. The typical tendency of judgment (that vigorous lifters give more "rights" with J <S than with S> J), finally, which proposition 5 is designed to explain, may be taken simply as showing that vigorous individuals develop greater precision in estimating a decrease and smaller in detecting an increase of stimulation. It is of course almost always true that the degree of precision (h) is not identical for heavier and lighter psychometric functions with the same standard. An increase of weight evokes criteria of one form or type, a decrease those of another; so that one group may well be more stable and sharply defined than the other.

(c) We have stated that Müller's concept is not explanatory; how may we then account for the preponderance of correct responses when J succeeds S? We may approach the question in two ways: (i) Direction of attention. Müller himself remarks the influence of O's attention upon a.i.; as Klein's observer, he gave special heed to the first impression of each pair and thereby wholly abolished the general judgment-tendency (19, 120). Klein

had an analogous propensity for attending to the right stimulus and so judged more correctly when J was to the right than to the left of S. When, as usual, judgment applies to the second stimulus, most observers seem disposed to give it more attention as the final and crucial impression. Even were S, when it precedes, to be wholly unnoticed, attention to the following J would serve to set off its criteria from those of the oft-experienced residual impressions of S and thereby induce some right responses; but no amount of attending to S when second will give any clue to the preceding J, which may be heavier or lighter. Time-order S-J therefore yields more rights than does the opposed order.

(ii) To the above approach by way of the conventional concept of attention we may add another in terms of the gestalt-doctrine. It is a commonplace of perception that a new impression tends to be "assimilated" to a frequent or usual one. When J precedes, it will tend to arouse not merely its own characteristic impression but the S-process as well, the two having repeatedly combined in a single configuration before; by the time S arrives, therefore, the whole SJ gestalt with its criteria of disparity is already sub-aroused. If this sub-arousal be sufficiently intensive, J will appear "absolutely" heavy or light before S is given; because the same SJ pattern has previously yielded this same response (heavier). When the sub-arousal is fainter but still present, the addition of S a few moments later will add but little to the existing energy-interchange within the pattern; whence the disparity or e.m.f. between the two neural excitations is lessened and discrimination becomes difficult. When S precedes, which has repeatedly been configured with every one of the J, it tends to activate all the SJ patterns equally and therefore not revive any of them to an appreciable degree; accordingly, when J is applied, a sharp disparity or p.d. results and discrimination is made easy. The time-orders differ essentially in the rate of energy-interchange within the SJ-gestalt; the faster this rate, that is, the greater the p.d. or e.m.f. within the pattern, the easier and finer discrimination proves to be.

(3) We now come to our final question: can absolute im-

pression and the judgment-tendencies be abolished? According to Müller a.i. can be done away if O be specially instructed to attend "hauptsächlich" and "ausdrücklich und in ganz besonderem Grade" to the first stimulus of each pair (19, 118). The end is much more simply and in our opinion reliably attained by a method used in the writer's study of thermal discrimination (7): using each stimulus in turn as the standard. Let there be seven weights, going from 85 to 115 grams by 5-gram steps. In one series I (85) is made standard and compared with all the others for say 12 trials; then II (90) becomes S for 12 trials and is compared with each of the others; and so on. We thus achieve the important end of having all the stimuli experienced in a given period the same number of times. Weight I is used for S 12 times and for J 12 times; so with all the others. No stimulus has the advantage of frequency over any other,[1] whence it is hard to see how any sort of judgment-tendency could possibly arise. Even were one to appear temporarily (say when I is standard), it would tend to arise equally with all the weights used, and the several tendencies would then compensate in the final combination of results. In our study referred to, the general tendency of judgment, that is the preponderance of "rights" in time-order S-J over the opposed order J-S, amount to 2.7 per cent; the typical tendency, that is, the preponderance of "rights" in group I (+D) over group II (—D), equaled 1.6 per cent. These figures well show how effective our method is in nullifying the Urteilstendenzen; whenever one desires therefore to eliminate instead of analyzing them, the above procedure may well be employed.

To sum up: we have tried to relate judgment by absolute impression to more simple and immediate forms of comparison; to show the correct method of finding the judgment-tendencies, to develop a theory of their origin, and finally to propose a simple and reliable means of preventing them.

[1] Unless more than one stimulator is given the same weight.

III. THE METHODOLOGICAL IMPORTANCE OF XI

We here desire briefly to notice the methodological value of Xi in psychometric work.

In the so-called psychometric function $\Phi(\gamma)$, we assume that as we increase the magnitude of a comparison-weight (J) in relation to some standard weight (S), the proportion of "heavier" (H) judgments will rise, not linearly as one might suppose, but at first slowly, then rapidly, then again slowly; the whole curve forming a gaussian ogive, the integral of a normal frequency distribution. In like manner the "lighters" (L) trace a corresponding ogive while the "doubtful" judgments distribute in some fashion between the two.[1]

Where now shall we set the boundary-line or point of indifference between H and L? There are several possibilities to consider:

(1) At the objective standard, S. In theory we should expect 100 gm to be rated H and L about equally often in comparison with 100 gm; S might therefore seem the natural or "true" division between H and L. But a set of comparisons is always affected with a variety of tendencies which are inherent in the experimental situation; and any limen derived under these conditions will inevitably have some errors imbedded in it. This objective standard therefore tells us nothing about the subjective line of demarcation between H and L in a given case; but it is a good criterion by which to determine whether and to what degree constant tendencies are at work. The less the subjective boundary departs from S, the more nearly have all constant errors been evaded or nullified.

(2) At the central tendency (mean, mode or median) of the "doubtfuls." No figure drawn from the distribution of these judgments can be considered a reliable or useful indicator of the

[1] The "doubtfuls" will be truly symmetrical only when the "heavier" and "lighter" curves both have the same precision or s.d.; which of course is rarely true.

boundary between H and L, for the reason that they have no direct relation to the curves which determine these two thresholds. Every comparison issues in one of the three responses, H, L or D; in percentages, D=100—(H+L). This equation of course reveals nothing about the proportion of H and L; the relative magnitude of the two may vary indefinitely without affecting D at all so long as their sum remains constant. An average based on D is then obviously no satisfactory means of fixing the point where " heavier " ends and " lighter " begins.

(3) At the arithmetic mean of the H and L limens. If the former is 3.0 and the latter is —3.2, their mean is —.1 which lies .1 gram below the standard. By this method however we assume that the two limens are equal in absolute value, for which there is no justification whatever. The limens for H and L may indeed prove equal in a given case, but there is no a priori reason why they should. The two series of judgments involve differing processes, H and L have unlike criteria; hence we cannot assume the two equal unless a wide margin of error is tolerable.

(4) Finally we may draw the subjective line between " heavier " and " lighter " at the place where the two psychometric functions intersect, where H and L are equally probable. Urban calls this place the " point of subjective equality " (25, xvi, 201). Being fixed by the two curves conjointly, it is immediately related with both limens and thus becomes the natural " zero " or indifference-point whence each limen is to be reckoned. The true H-limen then is the distance from Xi, where the percentage of " heavier " just equals the proportion of " lighter," to the median-point of the H-ogive, where $p_H = .50$.

We may illustrate the problem by an example from our thermal study (7). With an S of 24.00° " cooler " begins to appear at 24.15° which is far above the warm-median (24.086°), and " warmer " continues down to 23.95° which is well below the cold-median (23.988°). In a case of that kind where shall we say that " warmer " ends and " cooler " begins? The point Xi, where the tendencies to respond W and C precisely balance so that both are equally likely to appear, just fits our need; the true

limens, measured from Xi which falls at 24.047°, thus become .049° and —.049°; that they happen to be equal is of course merely coincidental.

By giving us the true point of subjective indifference, Xi enables us to detect and measure certain constant errors which other means do not even touch at all. Any limen as directly computed from the original data is affected with a combination of constant tendencies (time-, space-, movement-, and other errors). It is commonly assumed that upon combining the values for the four time-space orders we secure a mean limen freed of constant error. This is by no means true; other tendencies, notably habituation or expectation, still remain imbedded in the figures and cannot be removed by mere combination. This habituative or expective error, for example, arises as follows. Given an S of 100 gm and a set of five J symmetrically grouped about it, ranging say from 96 to 104 by 2-gram steps. Were no constant tendency of any sort involved we should expect the number of H and of L to be the same when all five J are compared with S an equal number of times[2]; but owing to various tendencies we may find a propensity to overestimate J and so get a disproportionate number of H. This preponderance of H may lead to a kind of habituation; that is, it may induce a tendency to say "heavier" in doubtful cases and thereby artificially decrease the H-limen. Despite every precaution this mode of error appeared in our thermal study. It was found that in *every* case the sign of Xi was opposed to that of p_w-p_c (percentage of "warmer" less percentage of "cooler"). At 16° e.g. the mean value of Xi for all time-space orders was —.006°; which means that quite apart from time-space errors "warmer" and "cooler" were equally probable, not when S and J were the same but when J was .006° lower than S. This fact shows that something was favoring the "warm" response; and therewith we find, sure enough, a predominance of W (+.02). In every case this opposed relation appears; when W is in excess Xi is negative; otherwise it is posi-

[2] This will be strictly true only when the two limens are identical in absolute amount.

tive. We see then that even when the proportion of "warmer" and "cooler" differed by only a few per cent, some degree of error was still induced.

How can we nullify this mode of error? The remedy which Boring prescribes (2, 282)—trying to keep the S-D symmetrical about the prospective limen, that is, about the place where preliminary observations lead us to expect the limen—is gravely inadequate; for general psychometric purposes indeed the proposal is irrelevant.[3] The best modern procedure (use of Vollreihen) disposes the J symmetrically about S, not about the prospective limen. We begin with a set of J equally spaced above and below S and then keep to the same set in determining both upper and lower limens (heavier and lighter); no guessing about the size of the limen is needed or indicated.[4] In brief, we hold it impossible, even under the simplest experimental conditions, so to dispose the stimuli in advance as to avoid a preponderance of H or L; the time-space errors alone, by favoring now one now the other, keep the two from being equally probable; and we have just seen that even when the frequencies of W and C differed by only a few per cent, the limens were in every case shifted to some degree in favor of the preponderant response. If the error appears with so small a disparity in the opposed judgments, we can never hope to remove it by arranging or adjusting the stimuli in advance;[5] the relative proportion of H and L or of W and C depends upon a complex of subjective (organic) tendencies whose influence can be measured and eliminated only *after* the event. Here is where Xi comes in. If H is in excess, its curve will be shifted downward and the crude H-limen will be too small (the crude L-limen being too large); but Xi will be correspondingly displaced and thus the true value of each limen can be determined. Even though the crude limens be affected with all sorts of error,

[3] Boring was apparently thinking of problems like the two-point threshold where this device is good as far as it goes.

[4] To save time and energy, we avoid S-D so large as to evoke all heavier or all lighter judgments; but this has nothing to do with the matter of habituation.

[5] In Urban's study (16, 287), for example, which was conducted with every precaution under conditions which may be considered ideal, the seven observers had a mean difference of 10 per cent in the frequencies of "heavier" and "lighter."

Xi provides an extraordinarily useful and convenient and reliable device for measuring the degree of their displacement.

The value of Xi and of the true limens can readily be found. Suppose the distance from the "warmer" median ($p_w = .50$) to the "cooler" median ($p_c = .50$) to be .116°, the gross warm limen (measured from S=32.00°) being .111°, the cold being .005°. Both figures are obviously much displaced by constant tendencies. The precision (h) of the "warmer" function is .304, of the "cooler" .379, their sum being .683. The true warm limen then equals (.379/.683) .116° =.065°, the cold limen being (.116°—.065°) or .051°. Xi finally equals (.111°—.065°)=.046°; that is, it lies at 32.046°.

For anyone using the customary psychometric functions, therefore, we believe Xi to be an indispensable tool.

IV. ON THE USE OF A VARIABLE STANDARD

In this article we wish to signalize the marked advantage of using a variable standard in psychometric work; that is, of using, within a given series, each comparison-stimulus (J) in turn as the standard. The discussion will cover three topics: description of method, possible objection, advantages.

First of all, what is the method? Given 10 weights, consisting of five S and five J, the latter ranging from 96 to 104 grams by 2-gm steps. Ordinarily the five S are all equal and midway of the J-series (100 gm); whence we have five D from —4 to +4. By our method the five S are duplicates of the five J, the two series being equal and parallel; we then have the same total equipment (10 stimulators) as before, but the number and range of D are much increased. Comparing the lowest S (96) with all the J, we get as D: 0, 2, 4, 6, 8; the next S (98) gives: —2, 0, 2, 4, 6; and so on. In tabular form we have:

S-D in gm.....	8	6	4	2	0	—2	—4	—6	—8
Orthodox method...			1	1	1	1	1		
Our method.....	1	2	3	4	5	4	3	2	1

Whereas by the one method we have but five D with a range of 8, by ours we have twenty-five D with a range of 16, symmetrically disposed about the standard. By the close of a working period every stimulator (both S and J) has been used precisely the same number of times.

But we start an objection at once: how is the limen affected by using each J in turn as the base? With a D of 2 grams, say, "heavier" will appear more often on base 96 than 100. Our method theoretically induces an error which, however, is neglible and can, if desired, be easily measured and removed. With the above group of stimuli, we have fifteen D from 0 up; 104 is the base for one D (0), 102 for two D (0 and 2), and so on. Now the Weber ratio for lifted weights may be taken as .025; hence we should expect a limen with base 104 to be about (4×.025)

larger than with the true standard, 100. Applying this reasoning to all the S-weights, we have:

$$\begin{array}{rr} 4 \times .025 \times 1 = & .100 \\ 2 \times .025 \times 2 = & .100 \\ 0 \times .025 \times 3 = & .000 \\ -2 \times .025 \times 4 = & -.200 \\ -4 \times .025 \times 5 = & -.500 \\ \hline 15 & -.500 \end{array}$$

Dividing by 15 we have —.033, which is the net error in this case; that is, the limen by our method is about .033 less than the true value (2.5) or 2.467; the two differ by 1.3 per cent. A disparity of this magnitude is small enough to be neglected; the margin of error from other sources will usually far exceed this; but of course we can easily multiply the limen as found by 1.013 and thus get the true value.[1]

What now are the advantages, the reasons for proposing our method at all in lieu of orthodox procedure? There are several: (1) By its means we achieve a maximal range and number of S-D with minimal equipment. In our example, if the zero-D are dispensable, all the others may be secured just as well from the J-series alone. This is a matter of moment when, as in our thermal study, the individual pieces are extremely bulky and expensive, so that each added unit makes the whole much more cumbrous and costly.

(2) The range of stimulus-differences, being just *twice* as wide as by the usual method, has the advantage of insuring some high and low values of p in every case; these, being notably more reliable, have likewise greater effect upon precision and limen both than does a p near .50.[2] We thus save time and energy by

[1] The net error is a function of three variables: the Weber ratio, the number of J, and the size of the steps between them. Thus, in our study of thermal discrimination the error in a typical case sank to about .0002°. We may also observe that in our calculation we assume each D to count equally in fixing the limen; this is not wholly true; the unlike percentages of "heavier" with D of 0, 2, 4, 6, 8 have unlike Urban weights and therefore unequal effect upon the position of the limen. The matter is not of much moment, but the D which preponderate below the true S have relatively low weight; so that our computation really tends to magnify what error creeps in by this source.

[2] Urban (25, xvi, 182) and Boring (2, 285) both contend that a p near .50 affects the position of the limen more than does any other; a plausible error, since the Urban weight, P, is maximal at p = .50 and minimal at p = .00 and

including some large S-D instead of confining ourselves to small D whose percentages (of heavier and lighter) vary little from .50.

(3) In addition to a few large S-D, the method provides more smaller ones as well (in the example, five D of O, four of +2, and so on). In view of (2) above we may be tempted to ask: Why not use nothing but large S-D, omitting the smaller ones wholly? All our p would then have high reliability with few or none around .50. To any such plan two facts are decisively opposed: (a) When we try to determine any theoretical function, sound policy invariably demands as many empirical measures as can reasonably be secured; the more ordinates we have the better. In lieu of the old method of two or three stimuli, we now use complete series of S-D (Vollreihen), which are expressly designed to give an empirical picture of differential perception through the range from 0 to 100 per cent. Percentages near .50 are then indispensable and can be made as reliable as in the extremes by simply taking more cases; which is just what our method does. (b) O's attitude and performance depend on the number of stimuli; his response to one is conditioned by others gone before; were there no small D, the large D would be otherwise perceived. Small D furthermore provide an incentive for the alert observer and give the comparative process a kind of "intellectual" inter-

1.00. Urban's P however is composed of two factors with wholly unlike function (art. VI): the one a purely mathematical device for reducing the data, the other an index of reliability, 1/pq. How much a given p will affect the course of a fitted curve is *wholly* due to the latter. This factor equals 4.0 at p = .50 but rises to 101.0 when p = .01 or .99; whence it follows that the Urban-function fits the tails much more neatly than the mid-range of a distribution. This over-weighting of the tail-percentages is necessitated by the greater stability of judgment near the extremes; it is but natural that these values define the smooth curve more than do the less reliable p around .50. Should the reader insist on empirical verification of the above statements, let him consider Boring's example (2, 284) where h = .3670 and L = .2001. How are these values affected if we omit in turn p = .16, p = .45 and p = .86, retaining in each case the other four? Here are the figures:

	h	L	Amount of change h	Amount of change L
True values	.3670	.2001		
Omitting p = .16	.410	.261	.043	.061
Omitting p = .45	.361	.180	.006	.020
Omitting p = .86	.321	.330	.046	.130

Omitting the p at .45 affects the value of both h and L far less than does omission of the other two; which proves of course that the p nearest .50 had but little to do with fixing the position of h and of L in the first place.

est; a few easy judgments are useful in keeping O " steady " and confident (2, 282), but if at all numerous he tends to become perfunctory in his work. In our method then we gain the benefits of a few easy comparisons and likewise have enough small D to make their frequencies reliable.

(4) A fourth advantage of our method is that we escape what Müller calls the effects of " absolute impression "; that is, the typical and general tendencies of judgment (19, 113). The reasons have been given in article II of this series and need not be repeated here. All the stimuli whether S or J being presented equally often, we fail to see how any trend, typical or general, could possibly arise; especially if we change S from trial to trial. In our thermal study, the general tendency reduced to 2.7 per cent, the typical to 1.6 per cent; these figures clearly show how effective our method is in nullifying the " judgment-tendencies."

For the reasons given we believe that the variable standard deserves careful attention from anyone seriously engaged in psychometric work.

V. ON THE URBAN WEIGHTS AND THEIR RELIABILITY

In this article we propose to examine the theory of the Urban weights, to demonstrate with an actual example their limited reliability, and to present a means of finding the true weight when needed.

Let S and J be two stimuli, presented to an observer under specified conditions; let O respond to J upon each presentation in one of three ways (Heavier, Doubtful, Lighter); let p_H be the proportion of Heavier out of the total number of judgments applying to a given J. As J is increased p_H will rise from .00 to 1.00, roughly describing in its ascent a curve known as the psychometric function or integral $\Phi(\gamma)$. Let a be the median of the ogive which best fits the ascending series of H-percentages and let h $(= 1/\sigma\sqrt{2})$ be the precision of their distribution. Then $\gamma_k = h(J_k - a)$. From a group of these equations we find the most probable values of h and a, such that the mean-square deviations of the observed p from the calculated p (p_c) will be minimal. Since the mean-square adjustment is designed only for directly observed quantities like p and not for derived values like γ, the γ-deviations are reduced to p-deviations by means of a factor, ψ, such that

$$\psi(\gamma - \gamma_0)^2 = (p - p_0)^2, \tag{1}$$

where p_0 is the most probable (adjusted) value of p and γ_0 is correlative with p_0. We have shown elsewhere[1] that

$$(p - p_0) = \frac{1}{\sqrt{\pi}}\left[e^{-\gamma^2}(\gamma - \gamma_0) + \gamma e^{-\gamma^2}(\gamma - \gamma_0)^2 - 1/3\,(1 - 2\gamma^2) e^{-\gamma^2}(\gamma - \gamma_0)^3 - \ldots \right]. \tag{2}$$

In general, the larger the number of terms in the right-hand member, the nearer will its value approach that of $(p - p_0)$.

[1] In our thermal study (7) to which the reader is referred for derivation of this formula and full discussion of the Urban method.

When employing the $\Phi(\gamma)$ procedure, however, we assume that our series of p_H trace an approximately Gaussian ogive. When this is true the magnitude of $(p - p_0)$ is small and likewise $(\gamma - \gamma_0)$ is so small that its higher powers may be neglected in the above series. Müller and Urban in fact retain only the first power of $(\gamma - \gamma_0)$, so that

$$(p-p_0)^2 = (1/\pi)\, e^{-2\gamma^2} (\gamma-\gamma_0)^2, \qquad (3)$$

and it is from this abbreviated equation that the weights in the Müller and Urban tables are derived. We see that the factors needed to minimize the sum of all the $(p - p_0)^2$ when we have the correlative figures for $(\gamma - \gamma_0)^2$ is

$$\psi = (1/\pi)\, e^{-2\gamma^2} \qquad (4)$$

But Urban in his weights (P) incorporates an additional factor of widely differing function: he weights each p in such wise that its contribution to the course of the psychometric curve will be proportional to its own precision or reliability. The p_H which we obtain with a given value of J is assumed to be itself merely a sample from a normal distribution; that is, if we made an indefinite number of trials at the same point (say a thousand sets of 50 each when J equals 102 and $S = 100$), the thousand values of p_H should form a closely normal distribution about their mean. The sequence of p_H (p_1, p_2, . . . , p_{1000}) evoked from a practised O under uniform conditions is taken to be a true Bernoullian series, the elements conditioning the appearance and nonappearance of a given response being constant throughout (cf. 10, 118 ff). The distribution has a s.d. of $\sqrt{pq/n}$, q being $(1 - p)$ and n the number of trials on which p is based (50 in our example). If n be constant, the s.d. of these p_H-distributions for the several values of J will obviously not be constant; for the reason that pq varies, being maximal at $p = q = 1/2$ and minimal when p or q equals zero. In fitting the most probable ogive to these values of p_H Urban weights each one, in the usual way, inversely proportional to its s.d.^{2}; that is each p weighs n/pq.

The Urban weight, P, therefore, unites two distinct functions:

(a) it enables us to apply the mean-square adjustment to our simple linear equations, $\gamma_k = h\ (J_k - a)$, so as to minimize $\Sigma\ (p - p_o)^2$ instead of $\Sigma\ (\gamma - \gamma_o)^2$: (b) it weights every p of the series to which the curve is being fitted inversely proportional to its own precision. Hence the course of the ogive is such that the sum of the weighted and squared deviations of p from p_o is minimal.

From the above it is clear that the complete Urban weight is (cf. equation 4):

$$P = \psi\ (n/pq); \qquad (5)$$

but for purposes of tabulation n is treated as constant (that is, unity) and ψ in each case is multiplied with $\pi/4$. This is quite permissible since only the *relative* weight of the several p is needed; while by so doing a more convenient table is secured, the maximal weight being reduced to unity instead of an improper fraction. In Urban's procedure, therefore, the following equation is fundamental:

$$(\pi/4pq)\ (p - p_o)^2 = P\ (\gamma - \gamma_o)^2. \qquad (6)$$

When the deviations $(\gamma - \gamma_o)$ are small, this equation is indeed accurate enough; but otherwise, an example from Urban's own work may be taken to show that it is far from correct.

Take the results for observer I, heavier (grösser) judgments, as given in the Archiv (25, xv, 287). The values of p_H for the several J (84, 88, – – –, 108) range from .0022 to .9400; that is, obs. I called 84 heavier than S (100 gm) in .0022 of the whole number of trials (450). The most probable values of h and a, calculated from these p's, are given (xvi, 192, Tabelle 44) as .1361 and 99.68. If we now compute the several p_o by inserting these values of h and a into the equation

$$p_o = 1/2 + \frac{1}{\sqrt{\pi}} \int_{\cdot}^{.1361\ (J - 99.68)} e^{-x^2} dx$$

and deduct them from the observed values p, the following residuals will be obtained in order: .0009, .0077, .0192, —.0172, —.1113, .0984, —.0054. The sum of these, when each is squared

and multiplied by the corresponding value of $\pi/4pq$, equals .129450. This is the figure for the left member of equation (6); the right member is computed as follows. The actual values of γ and the correlative weights P for obs. I are given in xvi, 190, Tabelle 40. If the several values of γ_0 be calculated from $h_0 (J - a_0)$ and deducted from the actual values of γ, these residuals will be obtained in order: .1190, .1376, .0924, —.0400, —.1985, .3008, —.0330. The sum of these, when squared and multiplied by the weights P, equals .098450 (Urban's figure is .098550; see xvi, 226, Tabelle 68). The agreement with .129450 is distinctly poor, and shows that the abbreviated equation with only the first power of $(\gamma - \gamma_0)$ is hardly satisfactory. If, however, the second-power term of equation (2) be retained in the weight, we have

$$(p-p_0)^2 = (1/\pi)\,[1 + \gamma\,(\gamma-\gamma_0)]^2\, e^{-2\gamma^2}\,(\gamma-\gamma_0)^2.$$

If the right member of equation (6) be calculated from this weight we get the value .1290 which, in view of the numerous approximations involved, may be called a good agreement with .1294.

The limited approximation of the weights in Urban's table could not be remedied merely by using additional terms from equation (2); by so doing we should introduce the variable $(\gamma - \gamma_0)$ into every weight; no table of reasonable size would then compass the possible combinations of p with $(\gamma - \gamma_0)$. So long as the deviations $(p - p_0)$ are small, we may well dispense with any finer approximation than the weights afford; but in the next article we shall have occasion to see, when considering Thomson's formula for the p.e. of a limen, how gravely the accuracy of his figures is prejudiced by the use of P which depart widely from their true value. In cases where the tabulated weight is not sufficiently precise, we may secure a much truer approximation by using additional terms in equation (2) above; as appeared so clearly in our example.[2]

[2] We trust this article may show once for all how needless and wasteful it is to retain four decimals in the Urban weights and to insist upon 100 trials (or a sub-multiple thereof) in a series for the purpose of securing integral

In general the Urban method will be used only with data which conform reasonably to the psychometric function; but the worker in this field will do well to bear clearly in mind the limited reliability of the tabulated Urban weights, and to make use of the more extended formulae when a truer value is required.

values of p, as many investigators still seem to do. The case herein examined (Urban's observer I) has no inversions of either first or second order and would commonly be regarded as a "good" set of data; and yet we have been able to show that the tabled Urban P are, in the mean, but 76 per cent of their true value. By the side of such a margin of error, an additional 1 or 2 per cent from dropping the third and fourth decimals is surely tolerable. In our own work we have never used more than two decimals either for p or for the tabled weights.

VI. THE PROBABLE ERROR OF THE LIMEN AND ITS DERIVATION

In lieu of Urban's and Thomson's complex formulae for the probable error of the limen [that is, of the median of the integral function $\Phi(\gamma)$], we wish here to propose a simple and equally reliable method of deriving and computing it.

Urban in his monograph computed, by a formula of his own, the p.e. of the upper and lower limens and of the intervals between them—the "intervals of uncertainty"—and obtained values which in every case were larger, by as much as 10 times, than the quantities whose deviations they represented (25, xvi, 225 ff and 192). These surprising figures lead him to conclude with the statement: "Die sogenannte Konstanzmethode ist demnach die ungenaueste der hier dargelegten vier psychophysischen Methoden" (p. 227). He ascribes its inaccuracy to the fact that errors of theory combine with errors of observation in the $\Phi(\gamma)$ method. That his explanation is forced and his derivation erroneous, however, is clear for several reasons: (1) It would be strange if data secured under optimal conditions from an observer as skilled as Urban himself, should have deviations so large as almost to deprive them of any value. (2) He shows himself (p. 224 f.) that the $\Phi(\gamma)$ curve agrees better with the observed data than does the arc-tan hypothesis, the other psychometric function which he tested; if this be true, how can the former be less accurate? (3) Since the 450 trials were made in nine groups of 50, each having its own value of h and a, the p.e. from these nine should roughly agree with the one derived from his formula. When so computed the p.e. of the limen equals .276, which is of a wholly different plane of magnitude than Urban's figure, 9.477 (cf. Thomson, 23, 52).

In lieu of Urban's formula for the p.e. Thomson, in the paper just cited, has developed another by the customary basic procedure:

If Z is any function, f, of $z_1, z_2, \ldots$ and if X is the error pro-

duced in Z by the concurrence of errors x_1 in z_1, x_2 in z_2, . . . , then

$$Z + X = f\,[(z_1 + x_1)\ (z_2 + x_2)\ \ldots].$$

Expanding the second term into a Taylor series in powers of $x_1, x_2, \ldots$, and deducting $Z = f\ (z_1, z_2, \ldots,)$, we have

$$X = x_1 \frac{\partial f}{\partial z_1} + x_2 \frac{\partial f}{\partial z_2} + \cdots + \frac{1}{2!}\left(x_1^2 \frac{\partial^2 f}{\partial z_1^2} + 2\,x_1 x_2 \frac{\partial^2 f}{\partial z_1 z_2} + x_2^2 \frac{\partial^2 f}{\partial z_2^2}\right) + \cdots$$

Assuming that the deviations are small enough to neglect all powers of $x_1, x_2, \ldots$, beyond the first, we have the usual formula

$$X = x_1 \frac{\partial f}{\partial z_1} + x_2 \frac{\partial f}{\partial z_2}$$

from which

$$p.e._Z^2 = \left(\frac{\partial f}{\partial z_1}\right)^2 p.e._{z_1}^2 + \left(\frac{\partial f}{\partial z_2}\right)^2 p.e._{z_2}^2 + \cdots$$

This formula is acceptably approximate only when the p.e. of $z_1, z_2, \ldots$ (that is, of $p_1, p_2, \ldots$) are quite small. The p.e. (p) is found by the Bernoullian formula $.6745\sqrt{pq/n}$.

The final form of Thomson's equation is given (23, 50). While presenting no inherent difficulties either in derivation or in application, the formula is so cumbrous that its use in computation becomes excessively laborious. It well behooves us therefore to look for a method that is equally reliable and much less toilsome and time-consuming.

Now it is obvious that the p.e. of both h $\left(=\frac{1}{\sigma\sqrt{2}}\right)$ and a (the limen) may be found directly from h itself. This quantity is the " measure of precision " of the differential (normal) curve whose integral is fitted to the observed percentages ($p_1, p_2, \ldots$). The s.d. of this normal distribution is $1/(h\sqrt{2})$, whence the p.e. of the median, a, is $.84535\ [1/(h\sqrt{2n})]$ and of h itself is $.6745\ (h/\sqrt{2n})$. For comparative purposes we shall deal with the same data (taken from Urban) which Thomson discusses, in which h=.1361, n=450; upon applying our equations above, we find that p.e.(h)=.003 (Thomson's figure=.003) and p.e.(a)=.207 (Thomson's value=.151).[1]

[1] Boring (3, 315) uses the same method in finding the p.e. but his derivation has two errors: (1) he determines the p.e. of the mean whereas the Urban limen is the *median*, of which the p.e. is greater than of the mean by .8454/.6745; (2) he uses a weighted value of *n* in lieu of the actual figure; thus from five series of 50 each (p. 317) he has five p of .40, .50, .60, .80, 1.00 whose correlative Urban weights (P) are .98, 1.00, .98, .77, .00. Each n being weighted with the corresponding P their sum becomes 186.50 and their mean 37.3, which he takes as the value of n (in lieu of 50) for computing the p.e. of the limen. We now have the anomalous result that, while a p of 1.00 (P = .00) has no effect whatever upon the position or variability of the limen, it grossly affects the calculated p.e.; indeed by using ever larger stimulus-differences and thereby getting more and more p at 1.00 we could enlarge the p.e. without limit. The method is clearly defective, and for this reason. We have above just five p of unlike weight from which to determine the psychometric curve; the absolute size of these weights (P) is wholly immaterial, their relative magnitude alone being significant. Of the five P four are larger than their mean (.746), the fifth very much smaller. The p of weight 1.00 has an *effective* value of 1.00/.746 or 1.34; likewise .98 of 1.31, .77 of 1.03 and .00 of .00; the relative contribution of the several p to the psychometric curve is shown by these figures. If we wish to weight the various n, we must then proceed as follows:

$$50 \times 1.34,\ 50 \times 1.31, \text{ and so on.}$$

The difference between these last two (the most important) figures is pretty large. We have now to choose between the Thomson formula, which gives an approximate p.e. of the limen for the exact (observed) values of p and the h-method, which gives an exact p.e. for the approximate (adjusted) values of p. Neither of course is wholly accurate: our h-method, because it shows the deviation for the "most probable" smooth (normal) curve instead of for the empirical values of p; and Thomson's, for at least two reasons: (a) because of the abbreviated p.e.-formula which he uses; and (b), what is doubtless more important, because of the merely approximate values of P (the Urban weights) which appear again and again (some 11 times) in his equation.[2] In the article just preceding we saw how much too small they are in this very case, being on the whole merely 76 per cent of their true value (.09845 as compared with .12945). The smallness of Thomson's figure (just 73 per cent of our own) seems undeniably due, in large if not major degree, to the fact that he uses a set of P which are about three-fourths of their true value and so gets a p.e. which is too small by about the same amount. Our h-method, we may add, is fully confirmed in this case by Wirth's interpolation-procedure (28, 192, where is found the original derivation). The Wirthian formula makes use of the extreme value, where $p = 1.00$. The highest p for obs. I being .94, however, this value is not precisely known; but we may assume with

The mean of these five products will of course be 50; which means that the weighting may be omitted altogether. The only reason for using 1.00 as the maximal weight (P) in Urban's table is convenience or simplicity; the maximal figure might just as well be 5.00 or .03, so long as the others all had the same proportional values. But Boring's weighted frequency, n, is a direct function of the particular values tabled by Urban; so that if the maximal weight were say 2.00 instead of 1.00, the weighted n would turn out to be just twice as large as above (74.6 instead of 37.3). Being thus arbitrary and factitious in character, this "weighting" of n has no real meaning.

[2] In solving, to be sure, the positive and negative values of P tend partially to nullify each other; their net contribution to the final result is thereby much reduced. From the nature of the formula we should judge the p.e. which it yields to be (roughly) about P times the figure which would emerge in case every P were simply dropped altogether from the equation. In the example (from Urban via Thomson) already cited, the tabulated P which Thomson employed in deriving the p.e. by his formula were, in the mean, only 76 per cent of their true values; in consequence his figure for the p.e. (.151) is but 73 per cent of ours (.207).

practical certainty, there being 94 per cent heavier judgments at 108 gm, that the ratio would rise to 100 per cent at either 112 or 116 gm, or at least approach 100 per cent so nearly as to make the error negligible. Setting 100 per cent at 112, we get a p.e. of .205, at 116 of .208; they both depart widely from the Thomson figure .151 while our .207 lies between the two. We may also note that our figure lies much nearer the rough check-value (.276) than does Thomson's.

In short our figure seems to be more accurate than Thomson's for at least three reasons: (a) he uses P which are known to be but 76 per cent of their true value and gets a p.e. which differs from ours by nearly the same amount; (b) Wirth's method clearly corroborates our figure; (c) our figure agrees better with the check-value drawn from the nine separate groups of 50 each.

Either method is really trustworthy only when the adjusted p differ but little from the observed. The evidence above presented all favors the h-procedure and we know of none which opposes it; and while we may not, without more ado, assume our method to be universally superior to the other, we may at least conclude that the Thomson procedure has no demonstrable advantage in its favor to outweigh the inordinate labor which its use involves[3]; for these reasons we used the h-method in our thermal study and commend it to the attention of other workers in the field.

One matter remains. Since both of our formulae involve n (the area of the normal curve fitted to the observed p), the question arises: what is the value of n when the number of trials upon which the several p are based is not constant? When each p is a certain proportion of the same number of judgments (450 in our example from Urban), then the area of the normal curve (n) is obviously this constant number (450). When however the number of trials with successive J is not the same (as by the

[3] One p.e. may be computed from h in about three minutes; by Thomson's method, if the writer may judge from his limited experience with it, something like four hours of intensive labor, with every mechanical aid for rapid work, are consumed in finding a single p.e. in a case like the above. When a great many have to be computed, as in our thermal study, this difference in time is of no little practical importance.

method which we recommend in Article IV), what area (n) shall we assume for the fitted curve? This is the formula:

$$n_m = \frac{\Sigma P_k n_k}{\Sigma P_k} \qquad (k = 1, 2, ---, s)$$

where n_m is the weighted mean value of n. In illustration of the procedure, take the following series of warmer judgments ($S = 20.00°$):

J	19.85°	19.90°	19.95°	20.00°	20.05°	20.10°	20.15°
p_w	.09	.29	.44	.51	.69	.69	.90
n	11	89	107	85	110	59	20
P	.51	.89	.99	1.00	.91	.91	.54

(J is the comparison-stimulus to which the judgment applies; p is the percentage of warmer out of the total number (n) of judgments taken for the corresponding value of J; P is the Urban weight.)

The sum of $P_k n_k$ here equals 440.55; the sum of P_k equals 5.75; n_m therefore equals 76.59. In other words, if 76.59 comparisons had been made in each of the seven series above, the p remaining as they now are, the whole psychometric curve would also have turned out precisely as it now is.

In cases where the Urban method is applicable at all, we believe the h-procedure above presented to be a reliable, simple and rapid way of computing the p.e. of the limen; and as such commend it to the notice of psychometric workers.[4]

[4] The reader will note that both the Thomson- and the h-method yield the p.e. of the crude (uncorrected) limen, which extends from S to that value of J where $p = .50$. The true limen, measured from Xi, will presumably, in our opinion, vary about the same; but inasmuch as we are acquainted with no procedure for finding its p.e., we have, for the present, to be content with the p.e. of the gross limen.

VII. ON THE PERCEPTION OF CHANGE

In this article we propose to examine the questions: What do we mean, in sense-perception, by an " intensive " and by a " qualitative " change? and, How do we come to perceive them?

No theory which has to do with comparison, the judgment of difference or the perception of change can evade this underlying problem; even though (dealing as we are with the ultimate nature of perceptual experience) we may be able to reply merely in suggestive or indicatory ways. We do not wish here to belabor again the hoary questions: whether an intensive change invariably conditions a qualitative alteration, whether " sensation " is a measurable magnitude, and the like; these have all been treated with an erudition that continues to be impressive even when it has ceased to be luminous. But having developed, in the course of our study on the thermal senses (7), a well-defined theory about the nature and criteria of sensory discrimination, we deem it well here to include a brief exposition of our point of view.

Let an observer immerse the hand successively in two jars of water; he thereby receives, within a brief interval, two complex somesthetic impressions, which as a rule differ perceptibly in certain ways; upon passing from one jar to the other, a noticeable change occurs. By this we mean that, insofar as the two are compared or otherwise related, they merge into a single gestalt, which gives rise to the perception of unlikeness and eventuates in some overt differential response (" warmer ") by the organism. Without making assumptions about the nature of this differentiative process or the character of the changes which appear, we may take them as given; together with the organic configurations which condition them, they are the ultimate source of any thermal discrimination. A similar process goes on and the same conceptions obviously apply with any other mode of sensory comparison as well. Two questions now arise:

(a) Could a naïve (untrained) observer tell of what *kind* or

nature a given difference is; whether "qualitative" or "intensive," say?

(b) Assuming (a) to be possible, could the same observer, if given a number of stimuli differing intensively, tell which is most and which is least intense (warmest, heaviest)?

We know of no reason for a positive answer to either question and therefore negative them both; in default of affirmative evidence we believe it wise to assume as little as we can, to wit: that a wholly naïve observer, with no antecedent opportunity of any sort for correlating subjective changes of impression (that is, the unlike and varied organic configurations) with objective changes of stimulation, could only report a "difference" or "change"; nothing more. He could observe that two successive stimuli were not identical but could not say whether a given change was qualitative or intensive, or whether an intensive change was upward or downward. The most primitive form of perception, the final source of all discrimination, we take to be this elemental *perception of difference or change,* which eventuates from the energy-interchange going on within a neural gestalt; but our untaught observer could not yet say whether a given gestalt was due to a qualitative or intensive change in stimulation, or whether to an increase or decrease in degree of stimulation.

To simplify exposition let us turn for a moment to the field of vision. Take the two series of qualities from red through orange to yellow and from red through light red to white. Each of the two has continuity, its members seem to belong together; our naïve O, if given samples chosen at random from the two, would presumably have no difficulty in restoring each sample to its own series and its true position therein. Question (a) now takes the form: is there anything whatever about these two continua that marks the transitions from R to W as a set of "intensive" changes and those from R to Y as a series of "qualitative" changes? In our opinion, there is nothing of the kind; two adjoining tints of red resemble each other no more and no less than do two hues of orange, if the difference in each case be equally perceptible; saturate red differs from Y no more than

it differs from a medium tint of red.[1] When the naïve O passes from one stimulus to another, he finds then no mark or sign to apprize him that two tints of red, say, or two warms differ in degree (more or less of the same thing) whereas R and Y or warm and cold differ in kind or quality. On the contrary, we come to look upon the transitions from red to white and from warm to warmer as "intensive" only because we have learned that they accompany changes in the intensity or degree of objective stimulation.

Question (b) remains: given two stimuli a and b from the intensive series (varying intensities of red light, or differing temperatures), could a naïve O tell which of the two is more intensive (lighter, warmer)? Again the reply is negative; we come to call b lighter or warmer than a for the sole and sufficient reason that we have previously found b to go with a higher degree of stimulation than a, or that the transition from a to b (that is, the configuration which results when they are presented together) reveals criteria which we have found to accompany a rising stimulus. What is our evidence for this view? It is twofold:

(1) *Objective.* The "all or none" law in muscular and nervous response, which "seems to have found favor with most of the workers who have investigated the matter" (15, 29), is pertinent here. An individual muscle-fiber when stimulated reacts with a maximal contraction if it gives any at all. The same law is found in neural function, according to the work of Adrian on the nerve-impulse, who proved "that a disturbance, after having been decreased by passing through a region of decrement, *recovers* its original magnitude when it reënters a normal area" (1, 385); whence Bayliss concludes that "there is no gradation of excitatory state in the normal condition" of nerve-fibers; a fact which "must be accepted whatever consequences may follow from it." This being true, we should not expect a perceptual pattern to show variations of degree or intensity, of more or

[1] We find in fact many more discriminable steps from R halfway up to W than from R to Y, as anyone can see by referring to Warren's color spindle (26, 171).

less within a specific cell; as the temperature rises and the response becomes "warmer," more nerve-units are actively involved in the total gestalt but there is no increase in excitation of any one. Warmer differs from warm therefore "extensively" rather than "intensively." [2]

(2) *Subjective.* This evidence is drawn from the writer's extensive experience in making some 35,000 temperature-discriminations and may be taken as the mature conviction which emerged during that protracted period of service. The testimony of others might not wholly agree with his; but the intensive and prolonged experience which it resumes may serve to justify its inclusion at this point as an important line of evidence.

"Assuming the second impression to rate warmer, is the difference of the two purely 'intensive'? So far as I can say, no; the second, to be sure, always resembles the first, but rarely if ever seems to be simply 'more of the same thing.' I am not sure I know just what this phrase means in actual experience; and this fact of itself may show how rare and unaccustomed the phenomenon was for me. In any event, the change from warm to warmer does not seem to be a mere matter of intensive increment. Typically, the second impression carries certain marks or characters which have been noted many times before upon passing from a lower to a higher temperature; the presence of these indices or criteria thus becomes the determinant of my response, and I call the present change warmer. Wherever it occurs, whether at 20°, 32° or 40°, this transition has a kind of family trait, bears much the same brand; somewhat as the notes of a melody continue to bear the same relation to each other even though pitched in various keys or played in differing parts of the scale." [3]

[2] We may think of the matter as follows: "warmer" is reported either when the first impression is cool and the second warm, or when each is warm and the second is more "intensive." In the former case, presumably, different receptors and therefore separate neural tracts are in action; in the latter, the second stimulus activates the same neurons as the first and some others in addition.

[3] The above statement is offered in evidence on question (2); and its brevity may show how hard it is, in any real sense, to give a significant description of the process of thermal discrimination. In a case of this kind, "introspection"

From these lines of evidence, however tentative and incomplete, we conclude that perceptual configurations (complex sense-impressions) vary in more ways than one; and that a given difference comes to be known as "intensive" solely for the reason that we have learned to ascribe it to a change in degree of objective stimulation. We may look at the matter somewhat as follows. We give O a number of preliminary series during which he is at all times apprised of the relation between the stimuli he is comparing; he knows, *e.g.,* in every case whether the second temperature is higher or lower than the first. If, under these conditions, we expose the same receptive mechanism again and again to a medium whose temperature (as known to O) is rising, we are likely to get a considerable variety of sensory configurations in the successive trials; some of the factors which constitute the several patterns are casual and incidental (arising from a particular jar, room-temperature, time of day, temporary physiological conditions, and the like) while others are stable or invariable, appearing every time. At each trial the judgment "warmer" is given; each gestalt, though differing in some way from the others which have preceded it, eventuates in the same overt response; and so the invariable elements common to all or most of the configurations, the modes wherein they are identical, become more intimately associated with, culminate more readily and surely in, the "warmer" response than do the occasional tingles, pricks, smoothness, lightness and other somesthetic factors which now appear in the gestalt and now do not. The fixation of these bonds between the stable elements of a perceptual configuration and some overt response (warmer) provides the *necessary* and *sufficient* condition for inducing true thermal judgments of "warmer." O's task in the training series therefore is to isolate

seems to be limited to suggestive and indicatory reports; by it we seem able to find out where something is going on and gain some clue to the nature and course of the ongoings, but the real mechanism of the process is "below the surface" and not discoverable by introspective means. One can, to be sure, describe the change from warm to warmer with metaphors borrowed from other senses ("cooler" is more open, light, smooth, velvety; "warmer" is more snug and prickly); but apart from the fact that we thus introduce irrelevant, non-thermal criteria which should have naught to do with O's judging, these figurative designations surely contribute nothing to our comprehension of the matter.

these invariable concomitants of thermal change until they stand in clear relief, and fix them by attentive repetition so well that future judgments will be consistently determined by them alone. This training is best achieved, not by haphazard " practice series " of the usual sort, but with systematic procedure wherein O is acquainted with the direction of thermal change and correlates it with concomitant variations in the perceptual configuration. With such tuition, O is gradually prepared to function properly in discriminative work, because he has gained a relatively clear and stable concept of " warmer " and " cooler." The improvement with practice which, as in our own case, may persist over many days of careful work, is therefore due not to changes in the receptive mechanism but to increasing refinement and clarity of the concept " warmer," which is gradually being isolated and defined.

Thermal[4] discrimination therefore consists in observing the kind of (subjective) criteria which appear within the configuration evoked by two thermal stimuli when applied under stated conditions to a given receptive surface. The totality of these indices and criteria makes up O's concept of " warmer " (and " cooler "); it is achieved by means of preliminary training wherein the stable features of the discriminative patterns come out with increasing precision and permanence, and are thereby relieved of casual and non-thermal adjuncts.

We have tried to show that all discrimination is in reality an estimate of stimulus-relations by means of certain subjective indices, which appear with neural or intra-organic configurations; that when O reports " warmer," he finds in the present gestalt those criteria which have previously gone with a transition from lower to higher temperature. By so doing, we may be accused of consorting, openly and unashamed, with the " stimulus-error." In a later article therefore we propose to submit this concept to a brief examination.

[4] We are, of course, merely using the perception of thermal change as an example; the same principles hold for any other sense-field.

VIII. ON THE USE OF AN INTERMEDIATE CATEGORY OF JUDGMENT ("EQUAL" OR "DOUBTFUL")

In this article we shall try to prove, by objective and experimental means, that with all comparative work involving the process of discrimination or the perception of relation (1) an intermediate category of judgment (such as "doubtful" or "equal") is indispensable for the most reliable and useful results; (2) "doubtful" is preferable to "equal" for this category.

The use and value of an intermediate response (in addition to the truly comparative judgments like heavier and lighter) has been much debated. Either "doubtful" (d) or "equal" (e), and often both, was used by the earlier psychometrists and by most of their successors; but Jastrow and others proposed using only the two comparative responses, of which one would always be "right" and the other "wrong" (provided the two stimuli are not equal). This is no place for an exhaustive history or analysis of the two methods[1]; but we may note some advantages claimed for the purely comparative procedure confined to the opposed responses heavier and lighter (the c-method):

(a) It simplifies manipulation and analysis of results; the d and e cases which were formerly troublesome are thereby wholly abolished (24, 286);

(b) It obviates d which is no real judgment at all, being a mere evasion or refusal by the observer, and not on a parity with true judgments like H and L[2];

(c) It precludes e which necessitates and issues from a kind

[1] See Titchener (24, 285), for a convenient summary of this topic's early history.

[2] Brown: "Evasive answers such as 'equal,' 'don't know,' or 'no difference,' permit the observer to avoid forming a judgment in just those cases where it is most important to discover whether or not a correct judgment can be formed" (6, 28); "When the difference is slight one must overcome some mental inertia before he can make up his mind, and if allowed to shirk, it is natural for one to lapse into a state of mild aboulia"; whereas by forcing O to choose between two alternatives we abolish the tendency to vacillate and keep him in a judicial attitude (p. 32).

of mental standard, which in turn is bound to vary from time to time and with different observers.[3]

The validity of these claims will be considered later.

Opposed to the c-method were considerations of this kind:

(a) If we compel O to "guess" when not sure, the cases which he would otherwise call d or e will now be assigned to H or L at random; but the law of chance-distribution (das Gesetz der Wahrscheinlichkeit) holds only with large numbers, whence a great many observations are needed if this procedure is to be permissible (Merkel: **18**, 586). It having been amply shown by Pierce and Jastrow (**20**) and by Brown (**6**) that O does not "guess" at random, this claim has no longer any force; even when professing to be wholly unsure of his judgment, O is more likely, apart from constant tendencies, to be "right" than "wrong."

(b) Suppose we have only positive D (stimulus-differences); in case O fails to perceive the increase (+ D) in a given trial he will incline to respond wrongly, "lighter" (**18**, 586). We are not quite sure of Merkel's meaning here; but anyhow, when we use the modern scientific procedure for systematic work (symmetrical distribution of D on each side of the standard, S) and measure the limen from Xi, as above advocated (Article V), any possibility of this kind is minimal if indeed operative at all.

(c) With small D where discrimination is difficult O will incline to observe with increased attention, in case d or e be denied him; and in consequence of this change of attitude will give proportionately too many "rights" (**18**, 586). Merkel tries to show this from Higier's results (**14**, 105). He calculates the value of h with Fechner's fundamental table for each D, both positive and negative, and finds that a D of —1 mm (S being 100 mm) gives much higher precision (h) in discrimination than a D of + 1 mm; this fact of itself, of course, merely proves the pres-

[3] Brown (**6**, 32): "The quality of sameness rarely or never occurs pure. It is necessary to decide in each instance *how near alike* the two things are and then decide whether that amounts to equality. In other words, one must maintain a mental standard of equality. As one becomes more or less scrupulous this standard is liable to vary; impressions that at one time seemed slightly different now pass for the same."

ence and operation of a constant error. But the mean of these two h for +1 per cent and —1 per cent is larger than of the two h for +2 per cent and —2 per cent and the rest of the higher D; he thereupon infers that "bei der kleinsten Zulage" O was observing with "verstärkter Aufmerksamkeit"; "möglicherweise wurde bei kleinen Reizen die Aufmerksamkeit überhaupt etwas höher angespannt als bei grossen" (p. 105). This conclusion is by no means warranted; all that Higier's data reveal is the operation of a constant tendency which naturally appears most noticeably with the smallest D; just as a displacement of .5 inch will alter the proportion of "rights" and "wrongs" much more when we compare small D of one inch than large D of ten inches. Nothing can be inferred from these data therefore except the presence of a constant tendency which appears more clearly with small D (±1 per cent) than with large D (±2 to ±5 per cent).

(d) A recurrent objection, finally, states that the simplified c-method straitjackets the observer and does violence to the comparative situation; O is not permitted to respond naturally but instead is brought under compulsion (Zwang) which distorts his natural mode of judging (cf. 24, 289 and 19, 14).

It is clear that previous discussion has been largely a priori, subjective or irrelevant; and inasmuch as the proponents of the c-method fail to accept the alleged introspective evidence against their procedure,[5] no agreement has yet been reached. Each investigator seems to follow his own preference in the matter; or if he wishes to use a particular method of statistical analysis (such as Urban's which, in its standard form, necessitates an intermediate category), he adopts the procedure it prescribes. This chaotic situation is regrettable; we therefore propose here

[5] Brown (6, 32): "The results of these experiments should be sufficient answer to . . . objections on the ground that the observer must not be forced or constrained out of the natural channels of expression. As a matter of fact no constraint is involved in the exclusion of these expressions ('equal' or 'don't know') in the case of observers who have had practice in attending to small differences. In the present case the observer gave this introspection in the middle of the minimal difference series: 'The weights never seem exactly the same. There is always a difference . . .' In the face of experience, a priori objections in this matter are out of place."

to present some objective and quantitative evidence in favor of the i-method (with intermediate category, d or e), which in our opinion proves it unquestionably superior to the other for general comparative purposes.[6] The argument will cover three heads.

(I) The i-method invariably has a higher degree of precision (h) or a smaller deviation for the psychometric function than does the c-method; it thus gives a more reliable and accurate measure of the discriminative capacity to be tested. We shall try to prove this with evidence from three independent sources. (a) From Brown himself. Let us compare all the experiments of Brown, four in number, by the c-method (including the various forms of instruction: " indicate which weight is the heavier " and " is the second heavy or light, in comparison with the first? " neither of which admits d or e) with two cases by the i-method where " same " was allowed (in one of which was the question: Is the second the same or heavier?; in the other: Are the two same or different?[7]). We here have materials for comparing those trials where O is forced to decide in terms of H or L and those where he may resort to " same " or " no difference."

[6] In this connection the reader may consult an article by George (13). He found that " doubtful " tends to lessen with practice whereas we found the reverse in our thermal study: at the close of practice " equal " had been wholly displaced by " doubtful." In our work O was instructed to announce " whether the second stimulus is warmer or cooler, that is, of higher or lower temperature, than the first." When they appear equal or indiscriminable, then " doubtful " is the true and only proper answer to the question. Our " doubtful " thus includes his " doubtful " and " equal " both; much of his attack on the " doubtful " response is thereby reft of its force. George invariably announced in advance that, say, " the second stimulus will always be greater than or equal to the first " and thereby directly suggested the use of " equal " where we should employ " doubtful." He concludes that doubtful judgments must " either be ruled out or discarded " (p. 11) and Boring (27, 444) insists that " the instructions and experimental setting must assure a practically complete elimination of doubtful judgments." In the present article we try to show that the doubtful category is highly useful even though we exclude all D-judgments from our final computations.

[7] In the former, O was apprised in advance that the second weight would never be less than the first, whence L was not an appropriate response. As we should expect, there were times when the second, despite this information, seemed lighter and was so reported; but, we are told (p. 30), these cases gradually ceased as the O tended to assimilate them to the " same " category. As a matter of technique, this procedure looks a bit curious but need not delay us here. In the second of these cases, we are told (p. 44), " the question was put, 'Are they the same (or nearly the same)?' and the judgment was again in oral terms 'same' but if they did not appear the same nothing was said."

To make the comparison easy and equitable we shall confine our survey to those D which appear in each of the six sets; the time-space order being the same in all cases (comparison-stimulus, J, on the left and second), conditions are strictly comparable except for slight variance of instruction and small discrepancies in the total number and range of D employed.

GROUP I

Percentage of "heavier" (p_H); J heavier, to left of S and second; S = 100 gm.

J > S by	2	4	6	8	10	12	14	16
1. (Which heavier? 19a + 23a)....	66	81	89	94	99	99	100	100
2. (Same or heavier?).....	50	69	86	91	97	98	100	100

GROUP II

Percentage of p_H; J lighter but estimated heavier, to left of S and second; S = 100 gm.

J < S by	1.6	1.4	1.2	1.0	.8	.6	.4	.2
3. (Which heavier? min. diff.).....	51	54	55	56	56	56	58	61

GROUP III

Percentage of "lighter" (p_L); J lighter, to left of S and second; S = 100 gm.

J < S by	2	4	6	8	10	12	14	16
4. (Which heavier? second less, motor).....	21	45	64	68	84	93	96	100
5. (Is second H or L? second less, oral).....	35	49	60	79	88	96	98	100
6. (Same or different? same).....	41	60	78	93	96	99	100	100

How are these figures obtained? Row 1: it is unfortunate for our purposes that Brown's successive percentages reveal large practice-effects; as appears when we consider the following series (pp. 14–15), which are identical in procedure being all listed as "greater, left." Three groups will serve for comparison: 1a to 5a, 17a + 19a, 23a (the last being separated from the next preceding series of the same kind, 19a, by some 30,000 trials; see p. 53); the precision and limen [8] for each of the three is:

	h	limen
1a to 5a	.210	2.27 gm
17a + 19a	.249	1.92
23a	.371	1.29

[8] The limen is assumed to extend from the point where $p_H = .50$ to $p_H = .75$; in other words, it equals the p.e. of the distribution $\left(\frac{.6745}{h\sqrt{2}}\right)$.

The precision (h) keeps on rising from early to later series, notably from second to third; this betokens a correlative improvement in sensitivity, wherefore reliable figures are not easy to secure for comparative purposes; but inasmuch as all the other observations here used were taken after 19a and before 23a, it seemed reasonable to assume that upon combining these two we should have about the same stage of practice represented in this method as in the others. Hence the figures given are the mean percentages for 19a and 23a. The figures of row 2 are copied from Table IV, p. 31; those of 3 from Table III, p. 24. In fitting our psychometric curve to the figures of 3 we of course bore in mind that the several D (.2 to 1.6 gm) were just 10 per cent of the D (2 to 16 gm) used in all the other series. The figures in 4 and 5 come from Table VII, p. 42; in 4 we have the percentage of "second less" as shown by a manual movement in accord with the instruction "indicate which weight is the heavier"; in 5 we have the proportion of "second less" as stated orally in reply to the question, "Is the second heavy or light (in comparison with the first)?" The figures in 6 finally are derived as follows: we form a cumulative curve from the "same" distribution (p. 42) by adding successively the number of "sames" in each of 200 trials, beginning at the right end (5, 6, 20, 40, and so on) and thereby get in order the sums 5, 11, 31, 71 and so on; the final sum, that is the total area of the "same" distribution, proves to be 1043. Dividing the sum at —2, at —4, and so on by 1043 we find the percentages in row 6, which reduce the "same" distribution to an ogive for comparison with the other ogival curves.[9] Altogether we have two modes of report (oral and motor) and two sets of D (+ and —) to compare with the series where "same" was admitted.

We now fit to these percentages the Urban psychometric func-

[9] The "same" distribution is unique in that it rises to a maximum and then recedes to zero; whereas all the others rise to a maximum (100 per cent) and then remain there. What we have really done here is to get the value of h (or σ) for the distribution of "same" judgments in a form which will make its precision (deviation) roughly comparable with that of the other curves.

tion $\Phi(\gamma)$ and get the value of h $\left(=\frac{1}{\sigma\sqrt{2}}\right)$ with its p.e. (see preceding Article VI of this series). h (or σ) is the only real measure of discriminative sensitivity and has the same meaning in every case; whereas "limen" is an arbitrary term which is variously defined in the two methods (c and i); we accordingly confine ourselves to the basal concept, h.

	h	p.e.$_h$	of trials
1. 19a + 23a	.288	.0137	100
2. Same or heavier?	.303	.0068	450
3. Minimal differences	.186	.0028	1000
4. Second less, motor	.292	.0098	200
5. Second less, oral	.291	.0098	200
6. Same	.382	.0129	200

The larger h is, the steeper is the psychometric curve and the smaller the p.e. of its distribution (the limen). We see at once that both cases of the i-method (3 and 6) show *higher* precision than do any of the other four; row 6 (.382) indeed exceeds the nearest c-value (.292) by about .090 which is 5½ times its own p.e. (.0162). The higher precision of the i-method appears the more sharply in that the two c-values with negative D (4 and 5) are almost identical[10] with each other and with the c-value for positive D (row 1); while all are substantially less than either i-value.[11]

We feel justified in concluding therefore that the two i-methods (same or heavier? same or different?) are *more precise* than any of the c-methods used; and while the differences in h are not large enough in every case to be valid by the usual criteria of significance (for the reason presumably that n is too small), yet the fact that both i-values are larger than any of the other four

[10] Brown seems to think (p. 53) that these two limens differ widely and then proceeds to draw rather far-reaching inferences about the general unreliability of the "threshold." On the contrary, the two methods agree almost perfectly in precision and consequently in differential limen.

[11] The figure for minimal differences (row 3) is much lower than any other and not so reliable; for the reason that the percentage curve has a kind of plateau in the middle and thus approximates the psychometric ogive but poorly. We cannot say which part of this curve most nearly follows its "true" course; but in any event the figure for h is at best approximative.

lends weight to the presumption that their differences are genuine. This being true, the i-methods prove to be the more precise and accurate mode of measuring differential responsiveness to weight-stimulation.

(b) To avoid the objection that Brown's experiments with the two methods (i and c) differed in other ways, however slight, as well and in consequence are not strictly comparable, let us consider the following data from Higier (14, 232 ff). These two sets of figures, on his formal statement, issue from identical conditions except that in one case the intermediate judgment (d) was admitted, in the other excluded. Those with d, it should be noted, were taken after those without d; he fails to state whether practice-improvement was provided for, but the figures show a little evidence of betterment from this source.

Percentage of judgments (longer or shorter); time-space factors constant; S = 100 mm.; n = 480

D =	−5	−4	−3	−2	−1	0	1	2	3	4	5
1. c-method, longer..	03	05	11	12	22	32	48	65	70	81	89
2. i-method, longer..	00	01	03	06	12	13	43	59	70	86	92
3. i-method, shorter.	96	91	87	76	61	41	18	15	08	03	02

These figures are from Table XIX, p. 277; the numbers there given are here reduced to percentages of 480. He tabulates the responses in the old way as "right" and "wrong," "longer" being right for a +D, "shorter" for a −D; but the values above are easily derived from his table.

Upon fitting the psychometric function to these three sets of figures, we have:

	h	p.e.$_h$
1. c-method	.227	.0049
2. i-method, longer	.308	.0067
3. i-method, shorter	.290	.0063

The differences with their p.e. are:

$$(1-2) = .081 \pm .0083$$
$$(1-3) = .063 \pm .0080$$

These differences are clearly significant, being eight to ten times their own p.e. Inasmuch as the experiments were made with identical conditions for the express purpose of comparing the two methods, we have to attribute a good deal of importance to these results.

(c) There remains the sole possibility that Higier failed to make due account of practice-improvement. To be reassured in this matter and to gain additional evidence, we proceeded to organize a brief experiment on lifted weights with rigorous conditions, the writer being observer. Actual experimentation was preceded with some hours of preliminary practice, during which O was in every case apprised of the actual relation between the stimuli; this was done with the purpose of defining and fixing the criteria for H and L. Five weights were used (194, 197, 200, 203, 206 gm.), with a rotating table to obviate the space-error. Time was marked by a clock beating seconds; each weight was lifted about one second with an interval of some four seconds between the members of a pair. In case of doubt, O "guessed" whether the second was H or L and then reported H-doubtful or L-doubtful; these cases were then tabulated in two ways: they were duly classified as H or L (c-method) and again as doubtful (i-method). By so doing we are enabled to use precisely the same data for our comparison; whenever any factor, like fatigue or practice, acts in any wise upon the one method, it affects the other as well; conditions are *really* identical as they can not be when observations by the two procedures are made at different times. Our data are therefore comparable down to the last detail, differing only in respect of method (the matter under investigation). The whole number of trials was 900 gained in six periods of 150 each. By using the variable S recommended above (Article V), we got nine D from —12 to +12, as follows:

D (in grams)	—12	—9	—6	—3	0	3	6	9	12
Relative no. of trials	1	2	3	4	5	4	3	2	1
Whole no. of trials	36	72	108	144	180	144	108	72	36
1. c-method, per cent heavier	17	19	34	38	49	56	63	69	81
2. i-method, per cent heavier	08	08	12	12	23	28	33	56	69
3. i-method, per cent lighter	64	61	40	33	20	19	08	14	08

Laying the most probable $\Phi(\gamma)$ curve through these figures, we get

	h	p.e.$_h$
1. c-method	.153	.0071
2. i-method, H	.174	.0082
3. i-method, L	—.180	.0085

As above noted, the observations were taken in six periods of 150 each; if we compute h at the close of each period, we have:

	c-method	i-method (H)	i-method (L)
First period	.142	.229	—.146
First two	.101	.149	—.126
First three	.122	.170	—.129
First four	.138	.167	—.156
First five	.161	.180	—.166
First six (as above)	.153	.174	—.180

While the figures fluctuate somewhat, notably in the first few days (owing to limited number of trials), and while h tends in all three to rise with practice, the relation is unmistakable: by the c-method h is *invariably* smaller than with i (heavier) or i (lighter); and so far as we can see there is no indication whatever in these figures that h in the first column will ever overtake h in the other two. Seeing no reason to continue, we therefore stopped observing at the close of the sixth period. The evidence from Brown and Higier is thus confirmed by our own experiment, designed and conducted with the express purpose of examining this matter.

In the face of such evidence we consider it proved beyond reasonable doubt that the use of an intermediate category of response insures greater precision and accuracy in measuring the differential responsiveness of the organism to a given mode of stimulation. This fact is directly opposed to the claim that the use of doubtful encourages O to shirk his task; but the evidence appears to us conclusive.

A little analysis will serve to show the theoretical plausibility of this conclusion. Consider Fig. 1. The comparative stimuli run from J_{-4} to J_{+4}, S being the standard. The ordinate in a given case shows the percentage of "heavier" (in a and b) and of "lighter" (in c and d). Let a be the $\Phi(\gamma)$ curve for "heavier" and d for "lighter" by the i-method; for convenience they are drawn symmetrical, though their precision in practice is rarely the same. Let b and c be the corresponding curves by the c-method; these are of necessity symmetrical, their precision being always identical. If no constant tendencies of any kind operate, both pairs of curves will intersect at S; with curves b

and c, H and L are equal at this point each being 50 per cent. It is clear that the dotted curves (c-method) rise more slowly, with lower precision and larger deviation than do a and d (i-method). The d will be most numerous in the region of S, while few will appear at the extremes, J_{+4} and J_{-4}; hence a

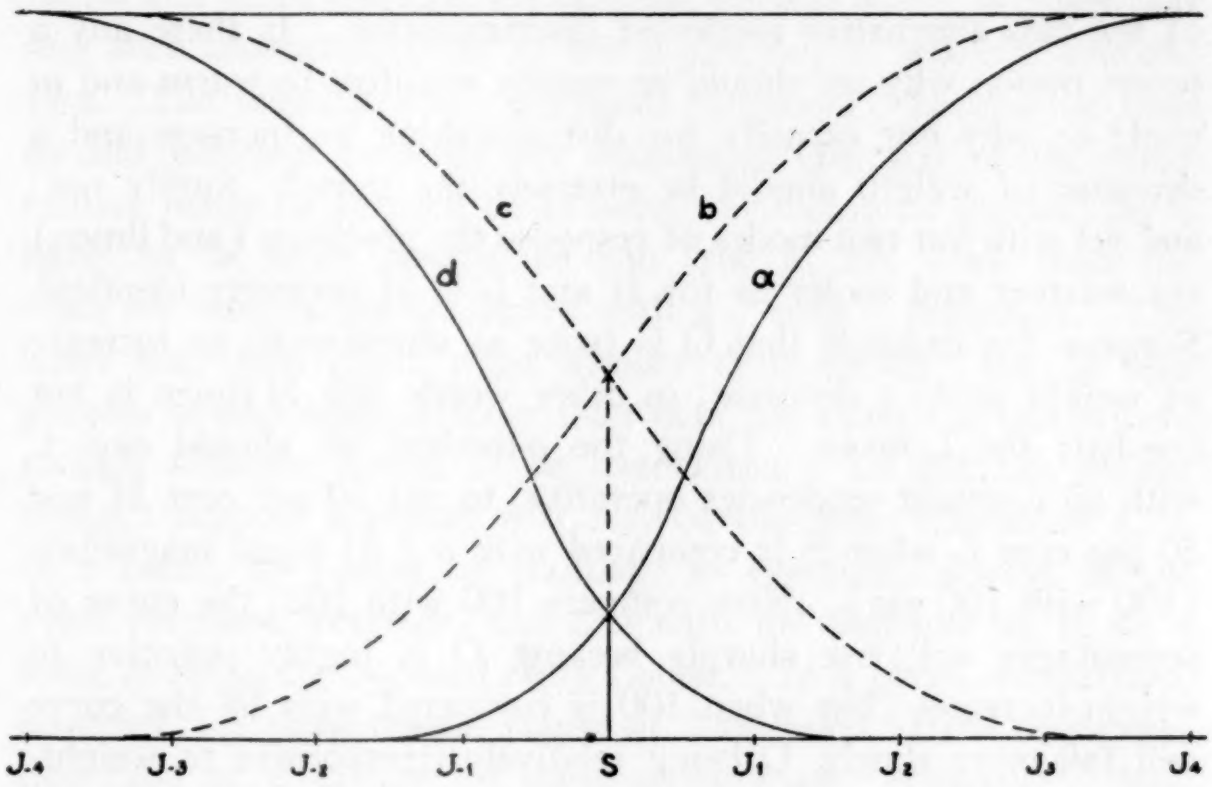

Fig. 1.

and b start near together at the top (with either method, c or i, we have about the same percentage of H at J_4) but diverge greatly at S where the large number of d depresses the a-curve much below b; around J_{-4} again the two curves are closely apposed. Hence if O reports heavier only when he is "sure" (that is, only when he finds those criteria which he has previously come to associate with a greater weight), we get a more precise measure of discrimination than when he is forced to "guess"; even though when guessing he is more likely to guess right than wrong. If it were a matter of getting the highest possible percentage of "rights" (as in a true-false examination), then we might well advise O always to guess when unsure; but when as here we are investigating his responsiveness to varying degrees

of stimulation, we get a more precise index of sensitivity when he judges only those D which meet the differential criteria he has established.

(II) A second disadvantage of the c-method is that the two psychometric functions (for H and L), being purely symmetrical, are compromises or hybrids which accurately represent neither of the two alternative modes of discrimination. Is there any a priori reason why we should be equally sensitive to warm and to cool? or why our capacity for distinguishing an increase and a decrease of weight should be precisely the same? Surely not; and yet with but two modes of response the precision (and limen) for warmer and cooler as for H and L is of necessity identical. Suppose for example that O is twice as sensitive to an increase of weight as to a decrease; in other words, the H-limen is but one-half the L-limen. Using the c-method we should expect, with no constant tendencies operative, to get 50 per cent H and 50 per cent L when S is compared with a J of equal magnitude (100 with 100 gm). Now compare 100 with 102; the curve of percentages will rise sharply because O is highly sensitive to weight-increases; but when 100 is compared with 98 the curve will fall more slowly, O being relatively irresponsive to weight-decreases. In other words the right half of the curve will rise with one rate, the left half will recede with a wholly different one; the two halves of the curve, instead of being symmetrical, belong to unlike psychometric functions; when we try to fit a smooth curve to such results we get a kind of hybrid function, whose precision lies somewhere between those of the two component halves. As soon as we admit the intermediate category d, this difficulty vanishes; the H-curve will then have high precision throughout, the L-curve likewise low; the two being independent, neither is compromised by the other. The d form a kind of neutral ground, such that the comparative judgments H and L can each trace its own course unprejudiced by the character of the other.

(III) The d-method, in our opinion, is more scientific and more exacting; it trains O to apply himself faithfully and induces

a larger measure of scientific caution; its responses are more precise because they proceed from a more scrupulous attitude.

These propositions are indeed the reverse of what proponents of the c-method claim: that the use of d promotes evasion and inertia whereas the comparative method keeps O from idle vacillation and from shirking his task (6, 28 and 32). We feel that any such characterization of the d-category is without foundation. The scientist who "suspends judgment" because he fails to find sufficient evidence for concluding one way or the other cannot justly be accused of evasion, vacillation or inertia; he is likely to be more scrupulous and true in his work than the observer who is always ready with a "snap" judgment. The latter may well guess right oftener than wrong; but he also guesses *wrong* much oftener than does the other—a fact which, though obvious, seems to be overlooked. To put the matter simply: suppose a gestalt with any of the three characters a, b, or c come, by course of training, to eventuate always in the overt response Warmer, and one with the features m, n, or o come to entail the opposed reply Cooler. In a given comparison, however, for some reason (lapse of "attention") the configuration fails to reveal any of these accustomed "earmarks"; how then shall O reply? Surely the scientific thing to do is to "suspend judgment" and report "doubtful"; that is precisely what we await from the seasoned man of science as opposed to the facile judgments of the untrained mind. Are not the same procedure and "spirit" of scientific caution equally useful and praiseworthy in psychometric service? If so, nothing is more fitted to destroy them than the habit of guessing. In our own experience with lifted weights, when forced to report H or L, we came to do so easily, with a shrug; forced to decide even when the criteria upon which we had come to depend were missing, we came to take the guess lightly, even flippantly; not as a last resort but as a commonplace. We are accordingly convinced that the guess is a much more fertile soil of perfunctory and heedless performance than is "doubtful," and that it is likely to compromise any form of scientific study or test.

In addition to being more scientific, the intermediate method is more exacting; it stimulates O to apply himself diligently and thus leads to uniform and efficient performance. When resorting to d, O fails to meet the primary task his instructions impose. He is required to report whether the second stimulus is H or L; his first duty then is so to clarify and extend his concept of H that a positive D, when applied, will evoke some of the characteristic features of this concept and thereby yield the correct response (H). Where this fails to occur, where the two stimuli condition a gestalt too vaguely or sketchily configured to entail one of the two comparative responses which he is primarily instructed to give, he is reduced to reporting d and thereby makes known that he is not equal in this trial to the task assigned. This is bound to be true if he makes the task of discrimination serious at all. Here is the chief psychological merit of d: being a *distinctive* mode of reply, d brings home unmistakably O's failure to meet the task imposed, and thereby insistently stimulates him to sustained attention and better mastery of his criteria. In the c-method, per contra, there is nothing distinctive about the guess-response; one L sounds the same as another; there is no repeated reminder that one's criteria are imperfect nor the same obvious incentive to revise and improve them; one may keep on guessing indefinitely and still meet the demands of the situation; improvement is thereby delayed and uncertain. We believe, for example, that Brown's observer found the i-series "more exacting" (p. 32) for one reason because the c-method had not constrained her properly to master and define her criteria. We saw (p. 101, above) that precision continued to increase after series 19a, which was already preceded by some 35,000 observations covering about five months; indeed the improvement seems to have continued well to the close of the whole set of 75,000 judgments. Practice-effects could hardly have persisted so long had the method been efficient. In our thermal study (7), improvement demonstrably ceased after some 7,500 observations within a period of 3–4 weeks; these figures look high and yet are only 10 to 20 per cent of Brown's. Thermal discrimination,

beset as it is with a multitude of confusing somesthetic impressions, is certainly no less complex and difficult than weight-discrimination; so that the proportionably slow improvement of Brown's observer amply confirms, we believe, our conclusions on the perfunctory character of the guess-method.

So much for the evidence in favor of an intermediate category; what of the objections raised above (p. 97)? The first, that it simplifies analysis, need not be taken seriously nowadays. The difficulties with d and e are obsolete; instead of apportioning them to H and L in some way or other, we simply neglect them; that is what they are for. They provide a neutral ground which permits each of the two comparative curves (H and L) to run its course unmodified by the other.

Secondly, is d an evasion or refusal by the observer? Far from it; if O is adequately trained for his duties at all, it is just as true a judgment as H or L, though differing in character from either. As above noted, d signifies absence of the stated criteria for H and L. In the preliminary trials of an experiment we isolate a group of criteria and then apply them as uniformly as we can: certain features in the whole impression touch off H, certain others L, while those complexes in which neither (or rarely both) appear are called d. It may well be that in these d-patterns are imbedded factors which may, and later do, come to serve as additional criteria of difference; as O improves with practice some configurations, formerly d, now eventuate in H or L; but so long as these factors remain obscure and undefined, the patterns in which they appear have no sure and uniform nexus with either of the two appropriate responses. Hence failure of given criteria to appear in a gestalt leads to d even as their presence entails H or L.

Thirdly, does an intermediate category demand some new kind of mental standard? None but the standards for H and L which any mode of discrimination presupposes. But will not the precision of the d-judgment vary from time to time and with different observers? It certainly will; even as does the precision of H, L and every other form of organic response. The per-

centage of d = 100 — (H + L); its frequency can change only as do theirs. So long as the standards for H and L remain constant, so will that of d; if we use those two how can we object to d which has a fixed relation to them? The frequency of d is fixed wholly and alone by our criteria for H and L; when the latter are complete and clear, the d will be few; otherwise numerous. Variation is just as true of the c-method as of the other. In either case O must find and stabilize his criteria as well as he can, but of course they will not be invariable. When O is dull and distracted, we may expect more "doubtfuls" than when he is fresh and attentive; but would anyone claim that O will "guess" equally well in these two conditions? We may indeed force O to give one of two alternatives in every case, but we cannot thereby force him to decide with equal precision every time.

We submit then that the objections to an intermediate category are obsolete, irrelevant or common to all forms of organic function.

Our second general question remains: which of the two modes of response is preferable for the intermediate category—equal or doubtful? The answer in turn depends upon another question: is either equal or doubtful an indispensable category, without which O cannot do justice to the psychological situation in which he is placed? Having agreed upon the use of an intermediate category, we must now adopt one that is adequate to O's every need.

In one part of our thermal study "equal" was imposed as the intermediate category between Warmer and Cooler, whereas in the method of constant stimuli "doubtful" was used throughout. We thus have materials for comparing the two. As for the former, the observers agree that not all "equals" are alike. One O reports at one time that she is "not doubtful about 'equal' judgments; they are as certain as the others"; she finds however that two impressions may differ in thermal quality, and yet she cannot be sure which is warmer. These do not really fit any of the three categories, but she calls them equal; doubtful would clearly be a more appropriate designation. Her repeated request

at the close of a series that, say, "the first two 'equals' be changed to 'warmer,'" shows clearly also that two impressions which at the time appeared equal did so no longer after other "equals" had been experienced. Another O finds: "Some 'equal' judgments are not very good ones. One 'equal' may seem better than the others—the recognition of equality is more certain. Some are not different enough to be called different but are not so good 'equals' as others." Later he says: "Equals more uniform to-day; no small differences between them."

From the above it is clear that if "equal" be admitted as one category, "doubtful" must also be used to accommodate the truly uncertain cases; for analytical purposes however the two can hardly be separated and are in fact usually combined; hence why use "equal" at all? The contention that "equal" may be just as certain as "warmer" or "cooler" may be true but is irrelevant; as we have shown in Article VII of the present series, the question for O in the process of comparison and the perception of relations is simply: Of these two sensory impressions, is the second warmer or cooler than the other? Or, in other words, of these two stimuli is the second of higher or lower temperature than the first? Or, once more, does the gestalt conditioned by these two stimuli have the characteristics which attend an increase or a decrease in temperature (from first to second)? When the two appear "equal" (indiscriminable) then "doubtful" is the *true* and *only* answer to these questions. The writer in his own observing therefore uses nothing but "doubtful" throughout; including thereunder the cases which reveal no apparent difference as well as those where the direction of difference is uncertain.

To sum up: we have shown by evidence from three independent sources that the use of an intermediate category of judgment invariably yields a psychometric function of higher precision (h) and smaller deviation (σ) than does the opposed procedure; it thereby gives a more reliable and precise measure of the differential capacities of the organism. We have shown, in the second place, that the c-method accurately represents neither of

the two alternative modes of discrimination (H and L) as does the i-method, but instead it gives a compromise curve fixed in varying degree by the criteria for H and for L. In the third place, we note that the d-method follows recognized scientific procedure and undertake to explain the proportionably slow improvement of Brown's observer as due to the perfunctory attitude which the guess-procedure encourages. The objections to an intermediate category we find to be obsolete, irrelevant or applicable in equal measure to every form of discriminative procedure. Finally, we note that "doubtful" is preferable to "equal" for this intermediate type of response.

In face of this evidence, we believe the superiority of the d-method has been established beyond reasonable doubt.

IX. ON ADOPTING THE PROBABLE ERROR OF THE PSYCHOMETRIC FUNCTION (URBAN) AS THE "LIMEN"

We here propose to set forth a few advantages of using the probable error (0.6745 σ) of the Urban psychometric function as the index or measure of discriminative capacity in lieu of the conventional "limen," which extends either from the objective standard to the median [the crude (Urban) limen] or preferably from the subjective standard, Xi, to the median [the "true" limen, as used in our thermal study (7)].[1]

The probable error of the Φ (γ) curve (hereafter called P) is in our view a more useful and reliable measure of discrimination than is the traditional limen (L) for at least four reasons:

(1) P has notably greater reliability than has the crude L, or median. The p.e. (L) is shown in Article VII to be

$$\text{p.e.}_{\text{Md}} = \frac{.84535\,\sigma_{\text{dis.}}}{\sqrt{n}} = \frac{1.253\,\text{p.e.}_{\text{dis.}}}{\sqrt{n}};$$

$$\text{and p.e.}_{\text{P}} = \frac{.6745\,\text{p.e.}_{\text{dis.}}}{\sqrt{2n}}.$$

Dividing second by first we have .381; in other words P is less than 2/5 as variable, or more than 2½ times as precise, as the median or Urban L. In both of course we postulate a normal frequency surface as does the whole Urban procedure. With so large a degree of difference P is undeniably superior in reliability.[2]

[1] Thomson (22) would debar the "doubtful" or "equal" response and then adopt the "interquartile range of the point of subjective equality," extending from 75 per cent heavier to 75 per cent lighter, as a measure of sensitivity. This proposal is defective in two ways: (a) by excluding "doubtful" he impairs the quality and precision of O's judgments (see VIII of this series); (b) he assumes the lighter and heavier limens to be equal, for which no experimental or theoretical justification exists.

[2] We have no method of finding the p.e. of the "true" L; the ordinary laws for the propagation of error fail to apply, since the factors which determine it are not independently variable; but we incline to believe that the true L varies less, or in any case no more, than does the Urban L.

(2) By adopting P we escape certain difficulties into which the use of "doubtful" (d) may occasionally lead. This intermediate category, as we have learned (VIII), serves to increase precision and thereby lower both the mean limen and its variability; but in case the observer has not adequately defined and stabilized his criteria, the following anomaly may from time to time appear. A decline in the proportion of d in the region of the subjective standard Xi has two opposed results: it lowers the precision (h) of the whole curve, thereby raising the limen and also appropinquates the median to Xi, thereby lowering the limen. For example, if O gave no d at all, the two comparative curves (heavier and lighter) would coincide at their 50 per cent points and both L reduce to zero. A reduction in d therefore tends both to raise and to lower L; of the two factors, the latter as a rule predominates. Two cases of independent origin will serve to clarify the matter.

(a) In our study of thermal discrimination, method B, 2 (constant stimulus-differences; simultaneous exposure of one hand to standard and of other to comparison; adaptation to 32°), we worked the same temperature (32°) twice, once at the very opening of practice and again after some months when the study was virtually done. In the first series O's procedure was not yet stabilized, wherefore the two disagree widely in the proportion of d; the percentages are shown in columns I and II.

Stimulus-differences		I	II
.25°		.00	
.20		.12	
.15		.46	.00
.10		.65	.17
.05		.78	.14
.00		.75	.28
—.05		.75	.32
—.10		.62	.44
—.15		.33	.03
—.20		.00	.19
—.25			.00
Warmer:	h	8.88	5.90
	L	.096	.048
	P	.057	.081
Colder:	h	—5.40	—2.64
	L	—.157	—.105
	P	—.088	—.181

The precision (h) in II is notably lower than in I, but the limen is smaller too. The situation is anomalous; when h is high, L should in all logic be low. The reason is not far to seek; the large proportion of d in the middle range of I enlarges both h and L, while the fewness of d in II reduces both L and h. This type of situation is of course unusual, indeed almost unique in our data; and merely shows that O in series I had not yet "settled down" to a definitive procedure. In passing we may note how clearly these figures confirm our proof (in VIII) of the value of the "doubtful" response in enhancing precision of judgment.[3]

(b) Another example may be drawn from Fernberger's study on the "stimulus-error" (9). His observers compared weights under three forms of instruction: pressure (A), kinesthesis (B), stimulus (C); the judgment in each case being greater (g), less (l) or equal. From observer B the following figures were obtained: the first three columns are taken from the table (p. 71); 4 and 5 give the true limens, each being inversely proportional to h; 6 and 7 contain the values of P.

	h_l	h_g	Interval ($=L_l+L_g$)	L_l	L_g	P_l	P_g
A.....	.111	.108	2.45	1.21	1.24	4.30	4.42
B.....	.112	.109	3.76	1.85	1.91	4.26	4.38
C.....	.112	.111	3.91	1.95	1.96	4.26	4.30

It appears that C has the highest mean precision and A the least; in spite of which C has the largest mean limen and A by far the smallest. Again we find that in each successive curve A, B, C (Chart II, p. 69) B increased the proportion of d thereby forcing the two medians ($p_l = p_g = .50$) apart, even while, in consequence of the freer use of d, she was raising her precision, that is, learning to differentiate smaller S-D. With the use of P as our discriminative index, no such anomalies can arise; as the percentage of d falls, h will likewise drop and P correspondingly enlarge, as in logic it should.

[3] When the frequency of d at the subjective standard of reference Xi, is below 50 per cent, it is clear that L will be less than P; with $d > .50$ it will exceed P, and when $d = .50$ the two will coincide. So here, in I where d goes up to .78, P is larger than L; in II where d stops at .44, P is less.

(3) A third value of the method is that it promotes a degree of "scientific caution" (Article VIII, 3) which the Urban procedure rather tends to inhibit. If O makes earnest with the discriminative process at all, he is sure to improve as the criteria come better in hand. Most good observers we believe tend to run ahead of their criteria and begin guessing when they should report d; the P-method reveals this procedure by raising the limen and so encourages O to wait for his criteria instead. We have presented (VIII) the value of the d-category in promoting scrupulosity; P goes a step further by promoting the use of d; it thereby trains O to depend on his criteria and in so doing to clarify and improve them. If we wish therefore a precise index proceeding from a scrupulous attitude, this revision of the $\Phi(\gamma)$ seems desirable.

(4) The method finally is simple and convenient, P being easily and quickly derived from h:

$$P = \text{p.e.}_{\text{dis.}} = \frac{.6745}{h\sqrt{2}} = \frac{.4769}{h}.$$

Its own variability in turn is exactly known:

$$\text{p.e.}(P) = \frac{.6745\ P}{\sqrt{2n}}.$$

While the deviation of the crude L is also known,

$$\text{p.e.}(L) = \text{p.e.}(Md) = \frac{.84535\ P}{\sqrt{n}},$$

that for the true L, a much more important figure, is still undetermined.

It may be of use to compare the true limen (L) with P when both are derived from the same data; we therefore present the following table (cf. 7, chap. VI), in which are shown both P and L for warmer and cooler by methods A,1 and A,2 (adaptation to temperature of standard in each case):

METHOD A, 1 (SUCCESSIVE PRESENTATION)

	28°	24°	20°	16°	20°	24°	40°	36°	32°	28°
L_w	.076	.064	.048	.067	.059	.039	.081	.062	.041	.038
p.e.	.013	.013	.010	.012	.010	.006	.040	.008	.006	.006
L_c	—.088	—.056	—.061	—.076	—.056	—.042	—.075	—.055	—.038	—.033
p.e.	.015	.011	.013	.013	.010	.007	.055	.008	.006	.005
P_w	.068	.083	.075	.086	.095	.060	.155	.069	.056	.058
p.e.	.004	.005	.004	.004	.004	.002	.012	.003	.002	.002
P_c	—.076	—.076	—.095	—.095	—.090	—.066	—.171	—.060	—.051	—.045
p.e.	.005	.004	.005	.005	.003	.003	.011	.003	.002	.002

METHOD A,2 (SIMULTANEOUS PRESENTATION)

L_w	.116	.095	.073	.077	.082	.065	.088	.068	.059
p.e.	.018	.018	.017	.032	.012	.016	.020	.013	.010
L_c	—.109	—.094	—.069	—.080	—.087	—.067	—.062	—.074	—.066
p.e.	.016	.020	.016	.030	.014	.016	.012	.015	.011
P_w	.104	.104	.108	.128	.123	.084	.142	.094	.070
p.e.	.007	.007	.006	.012	.005	.006	.008	.005	.003
P_c	—.092	—.102	—.099	—.132	—.132	—.087	—.096	—.104	—.082
p.e.	.006	.007	.006	.012	.005	.006	.005	.006	.004

WEIGHTED ADJUSTMENT EQUATIONS

(1) Values from 16°–24° ($y = a + bx$; $x = 0$ at 24°).
(2) Values from 32°–40° ($y = a + bx$; $x = 0$ at 32°).

METHOD A,1

	a	b		a	b
(1) 16°–24°					
L_w	.0392°	—.0035°	P_w	.0625°	—.0041°
L_c	—.0419	.0043	P_c	—.0684	.0045
L_w–L_c	.0811	—.0078	P_w–P_c	.1309	—.0086
(2) 32°–40°					
L_w	.0409	.0053	P_w	.0540	.0057
L_c	—.0383	—.0042	P_c	—.0484	—.0053
L_w–L_c	.0792	.0095	P_w–P_c	.1024	.0110

METHOD A,2

(1) 16°–24°					
L_w	.0677°	—.0023°	P_w	.0876°	—.0069°
L_c	—.0689	.0022	P_c	—.0902	.0064
L_w–L_c	.1366	.0045	P_w–P_c	.1778	.0133
(2) 32°–40°					
L_w	.0573	.0034	P_w	.0686	.0081
L_c	—.0685	.0005	P_c	—.0849	—.0019
L_w–L_c	.1258	.0029	P_w–P_c	.1535	.0100

In forming the normal equations for the fitted line ($y = a + bx$) by the usual method of least squares, each value (L or P) is weighted inversely to its $p.e.^2$.

The conclusions in our thermal study based on the limens (L) are, on the whole, confirmed by the P. (1) Practice-improvement again is found to cease at 20° (first series). (2) The P are

larger in absolute value than the L; this of itself has little moment, inasmuch as any index of discrimination has but relative significance. Here it proves merely that no stimulus-difference, whatever its magnitude, evoked as many as 50 per cent doubtfuls; in other words, the mode of the d-distribution, in the mean, was less than 50 per cent. (3) In the principal method (A,1) the Weber ratios are about the same, those for P running 10–15 per cent higher (.0086 vs. .0078 and .0110 vs. .0095). This may be ascribed largely if not wholly to the greater absolute magnitude of P; if each P be k times the corresponding L, it is clear that the Weber ratio (*i.e.*, the slope, b, of the fitted line $y = a + bx$), for P will also be k times that for L. It thus appears that an increase in the absolute limen carries a correlative but smaller increase in the Weber ratio or rate of rise in successive limens. When the minimal value of the threshold is high, its slope is also steeper. (4) The ratios for P in method A,2 depart widely from those of L and thereby confirm our previous conclusion (7, ch. VI): every important value of A,2 being but one-third to one-eighth as reliable as the correlative figure of A,1, the disparity in the Weber constants by the two methods may be due largely, or even wholly, to the high variability of A,2. Whereas with L the Weber ratio for A,2 is much smaller than for A,1, the reverse is true with P; in consequence the mean of the two ratios in A,2 is nearly the same as for the correlative values of A,1, and thus gives added proof that the figures of A,1 are reliable.

To sum up: we have aimed to show the superiority of the probable error of the psychometric function $\Phi(\gamma)$ over the conventional L as an index of discriminative sensitivity. Combined with Xi, the subjective point of reference, it provides everything the traditional method yields; and in addition it is far more reliable, it obviates the inconsistencies which may otherwise appear, it promotes a scrupulous attitude in O, while its magnitude and deviation can both be easily and exactly found. We comment it therefore to the use and criticism of psychometric workers.

Note on the comparative variability of the probable error and the quartile deviation in a normal surface.

We have just examined the advantage of using the probable error (P) or quartile deviation (Q) of the $\Phi(\gamma)$ as the measure of sensitivity. Inasmuch as the two are always identical in a normal surface it would seem to be irrelevant which term we use; when we come to apply the usual formulae of variability however we get widely divergent results:

$$\text{p.e.}(P) = (\text{p.e.}_{\text{dis.}}) = \frac{.6745^2\,\sigma}{\sqrt{2n}} = \frac{.3217\,\sigma}{\sqrt{n}}$$

$$\text{p.e.}(Q) = \frac{.5306\,\sigma}{\sqrt{n}}.$$

Q we see is 1.65 times as variable as P whence the latter is 2.72 $(= 1.65^2)$ times as reliable as Q.

We now have the anomalous result: of two values which are always identical in a normal curve one is much more reliable than the other. Even though they always coincide it becomes a matter of grave import which name we adopt when it comes to finding how reliable the value is. The relation as commonly stated cannot be strictly true; in 1,000 samples, if each were normally distributed, we should have 1,000 values for P and Q; the two being identical in magnitude, their variation would also have to be the same. That the difficulty is not merely personal or factitious was confirmed by two skilled statisticians, to whom the matter was mentioned; and inasmuch as the question is not considered in the common statistical treatises, the matter may be worth a brief note.

Take a normal surface where P and Q coincide. Now deduct from the lower half of the area an infinitesimal amount (smaller than any assignable quantity) and add it to the upper half. While P and Q will each of necessity be affected, Q will change in magnitude 1.65 times as much as P; at the instant of departure from normality, in other words, Q is 1.65 times as variable as P. In anormal surfaces P and Q have as many relations as there are differing modes of distribution; but as the frequencies ap-

proach normality by whatever route, p.e.(P) and p.e.(Q) approximate more and more nearly the limiting relation 3217/5306; and while this figure is commonly given as their relation in a normal curve, it is really true only when a curve approaches to or departs from normality. We must then have recourse to a fundamental concept of differential calculus, the limit, if we are to clarify what in its common form of statement appears absurd.

The true measure of deviation for our purposes is clearly p.e.(P) instead of p.e.(Q). h being the ultimate index of sensitivity by the method here proposed, it alone fixes the variability as well as the magnitude of the "limen," P. It follows that just as

$$P = .6745\,\sigma = \frac{.4769}{h},$$

so

$$\text{p.e.}(P) = \text{p.e.}(.6745\,\sigma) = \frac{.3217\,\sigma}{\sqrt{n}} = \frac{.2275}{h\sqrt{n}},$$

which is the value we used above.

X. THE STIMULUS-ERROR

In a preceding article (VII) we tried to show that the ordinary comparative judgment (heavier, warmer) is really an attempt to estimate the relation between two stimuli from the subjective indices which they evoke when presented together; that O, when properly prepared for his task, reports "warmer" because he finds in the present configuration those criteria which have in former trials gone with a transition from lower to higher temperature. By so doing we may invite the accusation of espousing the so-called "stimulus-error"; we here propose therefore to scrutinize this "error" and enquire into its nature and effects.[1] The argument will cover four topics:

1. What is the term meant to designate?
2. Is it a useful or apt designation for the purpose?
3. In what characteristic way, if any, does it operate?
4. Critical conclusions.

1. *What Does the Term Mean? What Is It Used to Designate?*

To reduce the question to elemental form, consider an observer who with stated control successively lifts two weights; he thereby receives two complex somesthetic impressions which as a rule differ perceptibly in certain ways; the whole configuration evoked by application of the stimulators eventuates in a differential response (heavier). In this perception of change will be certain features or parts (*Gestaltteile*) which continue to appear in successive trials and can thus be traced from one to the next, say T_1, T_2, T_3, S; the T being true criteria of weight (pressure, tendinous and muscular qualities, variously shaded, localized and combined), while S is a spurious criterion, irrelevant but confusing, like temperature (as the Weber illusion, whereby

[1] Boring (5) reviews the history of this concept in a paper which may be accepted as representative and about which our comment will center.

a cold object appears heavier than a warm one). All of these, alone or variously combined, may condition a given response. Likewise a comparison of temperatures will evoke a variety of patterns, in which may be found coolness now at fingertip, now between the fingers, now in skin-creases on back of hand and the like. These may all be called true criteria of cool; but the concomitant pricking, tingling, lightness, smoothness, snugness are secondary. The true indices commonly appear on a ground of irrelevant features which complicate the total impression and often become equally or even more prominent or "figural"; and which differ from true criteria in that they

(a) may appear only with certain degrees of stimulation; thus at 48°, where pain becomes insistent, discrimination changes markedly because thermal criteria are now being overshadowed or replaced by pain;

(b) are unlike the true criteria; thus cold water, because of its constrictive action on the dermal musculature, seems denser than warm; but this tactile quality is no true criterion for a thermal limen because it clearly does not resemble, or belong with, warm and cool;

(c) often vary independently of the stimulus; thus tingles and pricks appear at high and low temperatures both. In our records (7) was also noted several times the astonishing difference in tactual quality of two jars whose temperatures were almost identical.

Given a situation of this kind: we gather from Boring's discussion that an observer escapes the stimulus-error whenever his response is determined by one of these true criteria T alone (pressure on finger-tip, coolness on back of midfinger); but that O runs foul of the error in case his response (heavier)

(a) Is determined indeed by a single T but not by the same one each time, so that he shifts about from trial to trial, his report being conditioned now by T_1, next by T_2, then by T_3, and so on. His criteria then are genuine but vary from trial to trial (cf. 4, 447, and 5, 471).

(b) Is determined by S in whole or part; thus, in default of

thermal differentia, O may report " warmer " because his fingers sting (5, 463).

(c) Is determined by several T concomitantly ($T_1 + T_3$) instead of only one. This seems a possible, though by no means certain, implication of Fernberger's instruction; for if O fails to judge by pressure alone or by kinesthesis alone he is presumed to have the stimulus-attitude (9).

(d) Is determined by some combination of T and S, so that any change in the whole configuration, whether relevant or not, may touch off the response (4, 451).

(e) Is determined neither by T nor S directly but rather by the associations which they evoke; thus O's response might be affected by his visual representation of the weight's size, color, material, shape (5, 463).

(f) Is determined, finally, by no definite criterion of any sort, whether T or S, but issues at random in consequence of chance habits (4, 447).

We are by no means sure that all the above meanings are really sponsored by proponents of the term; but in any event this mode of listing may serve to clarify the question and to uncover the relation of the several possible meanings.

2. *Is the Term an Apt or Useful Designation?*

The use of a single criterion (T_k) throughout, which to Boring seems the test of good procedure, we heartily approve; but we do not therefore reject (a) and (c) as invalid. If in one trial coolness be noted only on the midfinger-tip and in another only by the line of immersion, are we not justified in reporting cooler both times even though the criteria be not identical? and if pressure on finger-tip and kinesthesis in forearm together induce the report heavier when neither alone would have evoked it, that judgment is surely permissible. If T_1 and T_2 are both valid criteria, how can T_1 plus T_2 or the series T_1, T_2, T_3, —, be other than sound and legitimate?

Type (e) furthermore is wholly acceptable in case the associations derive only from T and not from S. When comparing

sound-intensities, for example, O may image the height of fall in each trial; but in case the visual representation is conditioned wholly by T-factors and not at all by S, what is "erroneous" about so doing? Some may seek to avoid this procedure in the belief that it complicates the judging process; but we see nothing wrong or unsound about it.

The remaining three, in which the report is affected by irrelevant and specious factors, are of course erroneous. If O reports cooler on the basis of painful or tactile impressions, he is not doing what he is asked to do, he is not giving a thermal judgment at all. The complete absence of defined criteria (f), finally, can not occur under genuine experimental conditions; to attempt psychometric work without established criteria would be like professing to weigh without a unit of weight.

In three of the cases, therefore, neither part of the compound term "stimulus-error" seems to be relevant; the remaining three are indeed "errors" but seem to have no special relation to the "stimulus."[2]

3. *In What Characteristic Way, If Any, Does It Function? What Are Its Demonstrable Effects?*

Friedländer mentions three, of which only the last here concerns us:

(a) Verlagerung an den Aufmerksamkeitsort (11, 196); the objectified sense-impressions are referred to the region whither attention is directed.

(b) Verbreitung of the tactual sense-impression throughout the interior of the object as visually perceived or imaged, in such wise that the "density" of the sensory distribution varies inversely to the volume of the visual object (p. 197).

(c) Apparent change in differential sensitivity (limen), about

[2] Boring and others appear to use the terms stimulus and object as though synonymous, which in our view is an unfortunate identification of meanings. When I see a tree, the stimulus is a complicated pattern of light-waves impinging upon and activating the retinal receptors; but the object (tree) which I see is constituted by my own perceptual reaction. The stimulus is a form of energy, the object perceived is the organism's response to that stimulus. The stimulus exists apart from the perceiver, the object does not.

the direction and degree of which he professes to be still unsure (p. 187).

Of these three, quantitative findings, for which we are indebted to Friedländer, Fernberger and Reid, are available only for the third. (1) Friedländer had but one observer who was able to achieve and maintain the "stimulus-attitude," with the result that his data are confined to that individual. He employed two attitudes or Einstellungen: the object- or G-attitude (Gegenstandseinstellung), wherein the observer with eyes open is directed to compare the heaviness of the weights (Schwere der Gewichte); and the A-attitude (Abstraktionseinstellung), wherein observer with closed eyes is directed to attend to and compare the sense-impressions (primarily Druckempfindungen in the hand). He offers two sets of figures, for standards of 500 and 1200 gm respectively; but inasmuch as O completely failed, on Friedländer's formal statement, properly to differentiate the two attitudes in the latter case, the figures for 1200 gm. are without value for us. The other set (S = 500 gm), when treated by the Urban process, yields these values of h:[3]

	G	A
Lighter	—.0188	—.0168
Heavier	.0186	.0129

The p.e. of h being $\frac{.4769\,h}{\sqrt{n}}$, we have these values for the differences (h_G—h_A) with their p.e.:

Lighter	—.0020	±.00188
Heavier	.0057	±.00167

The latter difference, being 3.4 times its p.e., may be taken as fairly reliable; the other of course is not.

These figures being all derived in a single time-space order, we cannot find the true limen nor is it needful to do so; but the higher precision of G logically necessitates a lower threshold, as Friedländer also finds by his mode of reducing the data. Friedländer finds the lower limen of G "überraschend" (p. 191) for

[3] In order to make the values of h more nearly comparable we have included only those stimulus-differences which were used in both series (G and A), extending from 420 to 590, inclusive, by 20-gm intervals. S = 500 gm, n = 82.

the reason that O had trouble in maintaining the G-einstellung and so believed that his performance was better, more uniform and satisfactory, in the A-series. For an explanation of the fact Friedländer resorts to the concept of " sensation-density," that the sense-impression is spread throughout the volume of the object as visually represented. " Durch die bei der Objektivierung stattfindende Ausbreitung über das Volumen des Gegenstandes wird die Dichte der Empfindung vermindert, und wenn diese Abnahme der Dichte wie eine durch Reizschwächung hervorgerufene wirkte, so wäre eine Verfeinerung der Unterschiedsempfindlichkeit bei der Objektivierung zu erwarten " (p. 200). This is apparently what happens, even though " subjective conditions " in the G-attitude were " weit ungünstiger " than in A. Friedländer thinks the result may also be due to " grössere Anspannung der Aufmerksamkeit " induced by the very difficulty of G-Einstellung; his explanations are clearly tentative.

(2) In Fernberger's study lifted weights were compared with three forms of instruction: pressure (P), kinesthesis (K), stimulus-attitude (S), the judgment in each case being greater (g), less (l) or equal. Of the three observers, P had had " a great deal of practice," B " some little practice," H was " totally unpracticed." The following table gives the precision (h) for each observer by each method:

		h(l)	h(g)
P	P	.102	.103
	K	.131	.108
	S	.131	.117
B	P	.111	.108
	K	.112	.109
	S	.112	.111
H	P	.095	.097
	K	.093	.097
	S	.118	.112

The stimulus-attitude we see gives the highest mean precision in every case; so far the results conform with Friedländer's. This finding, in our opinion, is the quantitative solution of Fernberger's problem; neither he nor Boring, however, seems to think it worthy of special mention (p. 71 f., and 5, 465).

(3) Reid (21) followed Fernberger in using lifted weights

and the three types of instruction (pressure, kinesthesis, stimulus-attitude) with some refinements and extensions: more observers, more trials with each O, and at every session a report from O on his actual criteria during that series. He finds that no single O was able to follow each of the three instructions; they all exceeded or contaminated one or more of them (p. 73). Change of verbal instruction therefore does not of necessity induce a correlative change of attitude. An instruction further may be so difficult that O is unable to obey and perforce allows it to be contaminated. As for quantitative data, Reid gives the value of h for each instruction (P, K, S) for both judgments greater and less (42 values in all). Comparing h_s with the corresponding h_P and h_K for the seven observers, we find h_s larger in 21 out of 28 cases and smaller in 7.

The stimulus-attitude therefore yields higher precision, finer differentiation: with Friedländer in 2 out of 2 cases, with Fernberger in 10 out of 12 (the other two being equal), with Reid in 21 out of 28, the remaining 7 being smaller. The likelihood of this occurring by chance, if we may assume that the figures for P, K and S are truly independent (as they are if the three forms of instruction do not overlap and are rigidly observed; that is, if they are what they purport to be), is 1 in some 18 millions[4]; but of course neither of the two conditions is really met. Friedländer notes that, while the Druckempfindungen fade and recede under the G-einstellung, they still have a demonstrable influence upon discrimination and are by no means unregarded (11, 135 and 142); while Reid finds that no single O was able to keep every instruction pure (21, 73).

Be that as it may, the stimulus-attitude, within the conditions imposed by these three investigators, unmistakably excels in precision of result; it is clearly the method to use when we wish to test the differential capacities of the organism. As for explain-

[4] By chance, h in one method is just as likely to be larger as smaller than in another. Of the 42 cases cited, h_S is larger in 33, smaller in 7, equal in 2; the likelihood, in forty pairs of numbers drawn at random, that the first will exceed the second (or *vice versa*) 33 times is $\frac{40!}{33!\,7!}$.

ing this fact, we see just two possibilities: (a) The S-method may well be the most familiar. It is commonly said to be the "natural" attitude in the judgments of everyday life. If this be true, an O, when using his wonted mode of discrimination, might well show higher precision than when adopting a relatively new and unfamiliar one (judging by pressure or kinesthesis alone). No attempt was made, so far as we can see, to bring any O to his limit of practice in *all* three methods before undertaking comparative tests; Fernberger says they were given "sufficient preliminary practice so that the hand movements became automatic" (9, 67), while Reid states (21, 60) that "every O performed a practice-series of about 500 double lifts, which extended over a period of two weeks." Clearly if a subject who uses the right hand normally for writing were given some preliminary practice in manipulating the pencil with the left, we should not forthwith expect him to write as well with left hand as with right. Unless O be *equally* trained in all modes of discrimination, therefore, comparative figures prove little or nothing. Friedline (12, 415) claims indeed that "practice-effects," with the limen of dual discrimination, occur only with the "object-attitude" and not with the others; but Reid's data show improvement with practice in all three. If we compute the mean value of h in the first 250 trials and in the last 250 of both time-orders and all observers (except M and S who failed to follow instructions), and find the ratio of each pair, we have:

	Heavier	Lighter
S	1.21	1.20
P	1.12	1.05
K	1.10	1.09

The precision of the final series of 250 is greater than of the first by the ratios here shown; improvement thus appears with all three instructions but is twice as large for S as for the others. While not of major consequence then, practice can hardly be disregarded; and the precise import of Reid's and Fernberger's figures becomes hard to appraise.

(b) A more general explanation remains. Friedländer's observer had long training in comparing sensory impressions but

found it hard, as we saw, to maintain the stimulus-attitude. Whereas practice and subjective conditions seem all to have favored the former, the G-Einstellung none-the-less showed higher differential sensitivity than did the A. The only apparent reason seems to be that discrimination is improved by using a number of criteria (T_1, T_2, T_3, —), whether singly or collectively; if we have say three marks by which to distinguish two objects, we should expect discrimination to be more reliable and precise than when we have but one. The stimulus-attitude, by making use of *more* criteria, would thereby achieve a finer degree of differentiation. The same factor, we may presume, is likewise operative in Fernberger's and Reid's results.

The material of the present section in our opinion warrants the conclusion that the stimulus-attitude promotes discriminative efficiency and precision in the organism.

4. *Critical Conclusions*

We have tried to set down a few apparent meanings of the concept " stimulus-error " as found in current usage, and to see in what distinctive way its presence is revealed. In formulating a conclusion, our first problem is to analyze the psychological situation to which the term is applied. If we ask O once to judge the intensity of pressure sensations in the fingertips or die Druckempfindungen in der Hand zu vergleichen, and then ask him to judge the Schwere der Gewichte or the weights themselves, the two directions seem to be incommensurable; the second is hardly a psychological instruction at all of the same order as the first. To orient ourselves we must then put the stimulus-attitude into *psychological* terms and view it in relation to other modes of judging. For this end Friedländer's careful analysis may be helpful, who finds that two conditions are at once necessary and sufficient for objectifying the pressure and kinesthetic sensations produced by a resting or lifted weight: (1) die Richtung der Aufmerksamkeit auf den visuell wahrgenommenen oder vorgestellten Gegenstand; (2) the condition that " bereits eine gehäufte Zahl gleichartiger Wahrnehmungen vorhergegangen sein

muss, bei denen die Aufmerksamkeit dieselbe Richtung auf den Gegenstand hatte" (11, 193). The objectification of pressure and kinesthetic sensations, accordingly, is a product of experience, "die sich aus vielfältig wiederholten, bei gegenständlich gerichteter Aufmerksamkeit vollzogenen, gleichartigen Wahrnehmungen eines bestimmten Zusammenhanges zwischen visuellen und taktilen Sinnesinhalten bildet" (p. 201). All sense-impressions, he concludes, can thus be objectified.

Now we commit the stimulus-error, says Boring (5, 451) "if we base our psychological reports upon objects rather than upon the mental material itself, or if . . . we make judgments of the stimulus and not judgments of sensation." This antithetic characterization we feel to be neither sound in theory nor useful in practice. If the stimulus-attitude consists in attending to the object as visually *wahrgenommen* or *vorgestellt,* is the visual *Vorstellung* any less "mental" than the sensation to which it is opposed? It being clear, furthermore, that every comparative judgment, whatever name it be given (judgment of object or of mental material), can issue only from some kind of "mental" factors or criteria, whether well or ill defined, variable or constant, the antithesis gives no real clue to the psychological relation of the several methods which are being contrasted.

The matter in our view reduces to this: upon lifting two weights, we commonly image them in connection with the sensory criteria they provide (pressure, kinesthesis). This visual representation may affect the comparative process by supplying new differentia of its own, which are either absent or unnoted in the more immediate (sensory) criteria; somewhat analogous in effect though not in mechanism to the negative after-image, wherein one may observe features which one failed to remark in the positive impression.[5] A kind of conditioning process may sometimes occur:

[5] Within the ordinary conditions of psychometric work, one might question whether the visual representation brings any additional criteria to the comparative process, inasmuch as the stimulators are purposely made indistinguishable except by weight. But even here there may well be unconscious effects by these weight-differences upon the visual representation which O himself could not specify and which still may influence or aid his perception of difference.

originally the response "heavier" is evoked by a given pressure change (T_p) which is attended by a visual representation of the object; at first the latter may well be quite incidental and have nothing to do with the perception of difference; but in time it comes also to evoke the heavier response; whereas the pressure-factor, while always present and in a sense basal, may well become less obtrusive and clear. The visual *Vorstellung,* even though in the beginning wholly indifferent, may thus become a real adjuvant of the pressure and kinesthetic criteria of differentiation, so that in its absence the others fail to function as well as with its aid. This gradual conditioning process may well be responsible also for the more extensive practice-improvement in the stimulus-attitude.

As criteria for comparison therefore we may accept nothing but variations of pressure in the fingertips (T_p); or of kinesthetic impressions from the wrist (T_k); or finally we may accept any variation in the total impression, including T_p and T_k as well as any that may be provided by the visual representation (T_v).[6] When we are told about judging "objects" or "stimuli" we presume the third procedure is implied: judging from the total impression or pattern; so that A is reported heavier than B because of some relevant change in the whole configuration whether due to the direct and immediate action of present stimuli ("sensory") or to the revival of residual effects from earlier stimuli ("imaginal"). Using more criteria, the stimulus-attitude is both more complex and more precise. Being more complex, it is more beset with incidental variations and more liable to spurious criteria; many irrelevant changes go on which must be kept clear of the genuine changes due to the stimulus proper; but having isolated and clarified our criteria, we also have a more efficient instrument of differentiation than any single criterion can provide. These different modes of discrimination may well lead to unlike results, as noted above; but we thereby

[6] Somewhat as in the calculus, we may study how x varies when all other variables (y, z, . . .) are treated as constant; then find how y varies when x, z, . . . are held constant, and so on; finally we may treat them all as varying concomitantly.

see no reason for calling one of them an "insidious error," a "very real scientific devil" and similar epithets (5, 462, and 4, 449). To us they appear to be simply different modes of comparison, all equally valid in theory and all on the same plane methodologically, differing only in complexity (number and range of criteria). Certainly one may not use them all indiscriminately. Each has a function and value which none of the others may usurp; like any tool, each must be used in the way its character demands. But when used with skill and understanding, the stimulus-attitude is the most refined psychometric tool we have for measuring the functional capacities of the organism; and as such we believe deserves respect.[7]

The stimulus-attitude, we are told, "means indefiniteness and instability of criterion" and thereby necessitates equivocal correlation of stimulus with response (4, 447, and 5, 465). This may indeed be true if we take it up hastily and heedlessly; but the same may then be said of any psychological method. Being a refined and complex instrument, the stimulus-attitude needs greater care and skill for its proper use. We may briefly sketch how criteria should be isolated and controlled in lifting weights.

To acquaint O with the elemental phenomenon of change from heavy to equal to light and reverse, we give him a number of preliminary series in which he is at all times apprised of the actual relation between the stimuli (whether second is heavier or lighter than first). Little can here be done by verbal instruction; all we can do is to put O into a situation where heavy and light are juxtaposed and then let the criteria emerge and take on relief to which in future those terms (H and L) will be applied. In this way the crude criteria of everyday life are clarified and dissociated from specious and incidental material. This is specially needful in the somesthetic senses where a stimulator (metal weight) affects so many receptors at the same time and thus tends to evoke all manner of confusing configurations. Upon

[7] We incline to Friedländer's conclusion that A- and G-einstellung differ only in direction of attention; this makes the distinction all the more one of degree. When judging by pressure or kinesthesis alone (A-einstellung) the writer visualizes the fingers or wrist to which the pressure-patterns are referred, while in the stimulus-attitude he visualizes the weight itself.

lifting again and again pairs of weights whose relations are known to him, O gets a variety of sensory patterns in successive trials; some of the factors (Gestaltteile) are casual and incidental (owing to temporary physiological changes, variations in mode of seizing, raising and depositing weight, height and rate of lift, and so on), while others are stable appearing every time. With each positive increment, O reports "heavier"; each pattern, though differing in some way from those which have preceded, issues in the same overt response; and so the stable features come to eventuate more readily and surely in the report "heavier" than do the occasional variations which now appear and now do not. O's task in the training series is to isolate these invariable concomitants of change in weight until they stand in clear relief and to fix them so well that future judgments will be consistently determined by them alone. Under such tuition O is prepared really to function in discriminative work because he has gained a relatively clear and stable concept of "heavier" and "lighter." Any form of discrimination then consists in making a differential report based upon the organic or subjective criteria which appear within the configuration evoked by two stimuli when applied under stated conditions to a given receptive surface. The totality of these indices or criteria makes up O's concept of H and L at a given time; they are achieved by means of preliminary training wherein the stable features of discriminative patterns come out with increasing precision and permanence and are thereby cleared of casual and irrelevant adjuncts.

By this schema where does the stimulus-error come in? In Article VII we defended the view that all differential judgments are ultimately estimates of stimulus-relations; but of course they can only rest upon "subjective" or "intraorganic" criteria. Whatever we claim to be judging we are always judging by means of "mental" criteria and nothing else; and as we have just shown the stimulus-attitude yields to none in scrupulous attention to their formation and fixation.[8]

[8] Equivocal results on the limen of dual impression arise (5, 465) not because O adopts the stimulus-attitude, but because he fails to define his criteria.

Summary. We here try to set down various meanings of the term "stimulus-error," to consider its fitness for these meanings, and to find in what characteristic way its operation is revealed. We try to define the stimulus-attitude in psychological terms and to show that certain attributions, as indefiniteness and instability of criterion, are wholly unjustified. We seek to prove that it involves no scientific "error" of any kind; but that, on the contrary, it is, for certain purposes, our most refined psychometric tool.

References

1. Bayliss, W. H. Principles of General Physiology. Third ed., London, 1923.
2. Boring, E. G. Urban's Tables, etc. *Amer. J. of Psychol.*, 1917, **28**, 282.
3. Boring, E. G. The Number of Observations on Which a Limen May Be Based. *Amer. J. of Psychol.*, 1916, **27**, 315.
4. Boring, E. G. Control of Attitude in Psychophysical Experiments. *Psychol. Review*, 1920, **27**, 440.
5. Boring, E. G. The Stimulus-Error. *Amer. J. of Psychol.*, 1921, **32**, 459.
6. Brown, Warner. The Judgment of Difference, etc. Univ. Calif. Publ. in Psychol., 1910, **1**, No. 1.
7. Culler, Elmer. Thermal Discrimination and Weber's Law, with a Theory on the Nature and Function of Sensory Adaptation. *Arch. of Psychol*, 1926, No. 81.
8. Fechner, G. T. Elemente der Psychophysik. Bd. I, 3te Aufl., Leipzig, 1907.
9. Fernberger, S. W. Experimental Study of the "Stimulus Error." *Jour. Exper. Psychol.*, 1921, **4**, 63.
10. Fisher, Arne. Mathematical Theory of Probabilities. Second ed., New York, 1922.
11. Friedländer, H. Wahrnehmung der Schwere. *Zt. f. Psychol.*, 1919, **83**, I. Abt., 129.
12. Friedline, C. L. Discrimination of Cutaneous Patterns Below the Two-Point Limen. *Amer. J. of Psychol.*, 1918, **29**, 400.
13. George, S. S. The Effect of Instruction, etc. *Amer. J. of Psychol.*, 1917, **28**, 1.
14. Higier, H. Experimentelle Prüfung der psychophysischen Methoden im Bereiche des Raumsinnes der Netzhaut. *Phil. Stud.*, 1892, **7**, 232.
15. Howell, W. H. Textbook of Human Physiology. Eighth ed., Philadelphia, 1922.
16. Köhler, W. Die physischen Gestalten in Ruhe und im stationären Zustand. Braunschweig, 1920.
17. Martin, L. J., and Müller, G. E. Zur Analyse der Unterschiedsempfindlichkeit. Leipzig, 1899.
18. Merkel, J. Theoretische und experimentelle Begründung der Fehlermethoden. *Phil. Stud.*, 1892, **7**, 558; **8**, 97.
19. Müller, G. E. Gesichtspunkte und Tatsachen der psychophysischen Methodik. Wiesbaden, 1904.
20. Pierce, C. S., and Jastrow, J. On Small Differences of Sensation. *Mem. Nat. Acad. Sci.*, 1885, **3** (part 1), p. 75.
21. Reid, A. C. Effect of Varied Instruction on Perception of Lifted Weights. *Amer. J. of Psychol.*, 1924, **35**, 53.
22. Thomson, G. F. A New Point of View in the Interpretation of Threshold Measurements in Psychophysics. *Psychol. Rev.*, 1920, **27**, 300.

23. THOMSON, G. H. Accuracy of the Φ (γ) Process. *Brit. J. of Psychol.*, 1914, 7, 52.
24. TITCHENER, E. B. Experimental Psychology: Quantitative, Instructor's Manual. New York, 1905.
25. URBAN, F. M. Die psychophysischen Massmethoden als Grundlagen empirischer Messungen. *Arch. f. d. ges. Psychol.*, 1909, 15, 261; 16, 168.
26. WARREN, H. C. Human Psychology. Boston, 1919.
27. WASHBURN, M. F. Über den Einfluss der Gesichtsassoziationen auf die Raumwahrnehmungen der Haut. *Phil. Stud.*, 11.
28. WIRTH, W. Psychophysik. Leipzig, 1912.

ALLPORT'S EXPERIMENTS IN "SOCIAL FACILITATION"

By E. G. Williamson

Allport describes the social group as "any aggregate consisting of two or more persons who are assembled to perform some task, to deliberate upon some proposal or topic of interest, or to share some affective experience of common appeal."[1] In the "face-to-face" variety of such a group the members "react" to one another. In the "co-acting" variety the individuals are primarily occupied with some common "stimulus" beyond themselves; and there the "response" of the members to a common task or common experience is augmented by the sight and sound of others doing the same thing. Such an influence of the group upon its individual members is termed *social facilitation.*[2] In order to investigate the influence of this factor upon the members of a co-acting group and to compare the results with those of solitary workers, Allport had recourse to experiment. By keeping constant under both conditions such external details as furnishings, light, air and the seating of the subjects, the only remaining variables were, as he assumes, "rivalry" and the "facilitating" presence of others. These two factors produced "a distinct increase in the quantity and quality of the work of the individual." But rivalry was minimized or eliminated entirely, as he supposes, first by a constant amount of time given to each task and secondly by eliminating the comparison of achievements, with the understanding that the trials were in no way competitive. The subjects were instructed, however, to work at a maximal speed consistent with accuracy, and they were informed, moreover, that all were engaged upon the same task. This elimination of rivalry was designed to leave "social facilitation" as the single

[1] F. H. Allport, Social Psychology, 1924, 260.

[2] *Same,* Influence of the Group Upon Association and Thought, *J. Exper. Psychol.,* 1920, 3, 159–182.

variable factor producing an increase in the work of the individual while a member of a co-acting group.

So much for Allport; now we consider certain points in criticism. In the first place, it is not evident that the expedients named would really eliminate the factor of rivalry. A constant time for each trial may have prevented a race to finish first; but it is not clear that it would rule out competition in accuracy, in thoroughness, or in quantity of work. What was to prevent either the solitary or the grouped subject from competing against time or against a standard of accuracy whether set by himself or by the experimenter or implied in the setting of the task? This incentive or whip would be every bit as "social" as any rivalrous "desire to win." It is a social factor, moreover, which might appear quite differently in the isolated (A) and the grouped (T) work. A numerical difference appearing in the A and T performances might well be due to this factor as much as to "social facilitation." Allport ignores the fact that rivalry may functionate in a solitary individual as well as in a group-member. "Introspections" of the subjects, meager though they were,[3] suggest that the expedients did not as a matter of fact entirely eliminate rivalry.

In the second place, it was assumed that the only instruction which determined the performance of the subject was the verbal command of the experimenter. "Situation-instruction" and "self-instruction" were overlooked. As every descriptive psychologist knows, experience is determined from all these sources. The subject may instruct himself directly against the command of the experimenter, or the situational setting may nullify either.[4]

The third criticism has to do with the subjects. Information upon them is scanty. They were upper classmen and graduate

[3] Allport obviously understands "introspection" to include every sort of comment, reflection, speculation, whim or suggestion which occurs to the worker to report upon his work, his states, his opinions, the situation at large or the general run of events.

[4] FERNBERGER, S. W., Effect of Attitude of Subject Upon the Measure of Sensibility, *Amer. J. Psychol.*, 1914, **25**, 542; GEORGE, S. S., Attitude in Relation to Psychological Judgment, *Ibid.*, 1917, **28**, 1–37; BORING, E. G., The Stimulus-Error, *Ibid.*, 1921, **32**, 449–471; BENTLEY, M., The Field of Psychology, 1924, 390, and *Sup.* pp. 1–15.

students in psychology, twenty-six in number, although not more than fifteen were used in any one experiment. They were arranged in groups of three or five when working together. There is no indication that they were chosen with the purpose of eliminating those habituated by past experience to performance in groups. There was, furthermore, apparently no effort made to secure a fair sampling of the multifarious reactions of different individuals to social situations. Such factors might or might not influence the results; but it is better technique to control than to ignore them.

Now we turn to the numerical results from which Allport drew his inferences. In the *Journal* article, quantitative comparisons in terms of work done are made for " free chain association " and " thought process " (Tables II–X) with isolated and grouped workers (Experiments 1–4 and 6).[5] In Exper. 4 the average number of associations for the average subject is the same (27); in Exper. 6 the difference is negligible (8.0 ± 1.7 and 8.8 ∓ 1.9); in Exper. 2 the average excess for the grouped worker is 3.3 associations (with mean variations more than twice the differences), and in Exper. 3 about the same excess (but with m.v.'s of 21.0 and 20.1). Thus in only two of the four experiments does the associative output show a distinct difference favoring the work in the group, and there the high variability suggests that the small difference is without significance. Surely an insecure factual basis for the concept of social facilitation! But the case is still worse. A natural basis of comparison of the work of a given subject would be the total output (number of associations), provided the *number of trials was the same* in groups and in isolation. But Allport ignores the fact that most of his workers were subjected to more trials in the group than when alone.

In his Table II will be found a comparison of the number of trials for each subject working T and A at successive free associations. As reported the experiments are open to the criticism

[5] *J. Exper. Psychol.*, 1920, 3, 164–178.

that the numerical differences in the results T and A may be due to the difference in number of trials together and in isolation, instead of—as alleged—to "social facilitation." Let us compare the performance of the individual subjects on the basis of the *average number of associations per trial,* T and A. Our Table I is constructed from Allport's data on page 164 to show the results of such a comparison.

TABLE I

FREE CHAIN ASSOCIATION

Subj.	No. Trials		No. Assoc.		Aver. No. Assoc. per Trial		M.V. from Aver. of Aver.		
	A.	T.	A.	T.	A.	T.	A.	T.	
1.....	13	15	56	58.3	4.3	3.8	.9	.8	—
2.....	11	10	62.7	63.4	5.7	6.3	.5	1.7	+
3.....	5	6	67.8	78.7	13.5	13.1	8.3	8.5	—
4.....	11	10	54.3	54.4	4.8	5.4	.4	.8	+
5.....	6	6	44.2	48.3	7.3	8.0	2.1	3.4	+
6.....	13	16	56.8	55.1	4.3	3.4	.9	1.2	—
7.....	14	15	57.3	58.2	4.1	3.8	.11	.8	—
8.....	13	15	56.8	61.2	4.3	4.0	.9	.6	—
9.....	13	12	68.4	71.2	5.2	5.9	0	1.3	+
10.....	14	16	69.9	72.4	4.9	4.5	.3	.1	—
11.....	12	13	65.0	73.0	5.4	5.6	.2	1.0	+
12.....	14	15	67.4	67.5	4.8	4.4	.4	.2	—
13.....	14	16	40.1	46.5	2.8	2.9	2.4	1.7	+
14.....	13	12	76.9	80.9	5.8	6.7	.6	2.1	+
15.....	5	10	61.4	65.7	12.0	6.5	6.8	1.9	—
Average .	11.4	13.5	60.3	63.6	5.2	4.6	1.65	1.75	

The left half of the Table is Allport's; the right half is our own computation. In the final (right) column, "+" indicates performance favorable to the group, and "—" favorable to isolation. Where Allport found a difference of 3.3 (63.6–60.3) in favor of T, the difference is really 0.6 (5.2–4.6) in favor of isolation; though, to be sure, the m.v.'s indicate that this difference is not more significant than Allport's. But as the groups stand they are contrary to his interpretation. Since he counted noses, it is worth noting also that 8 *solitary* subjects had a higher average yield of associations, against 7 who stood ahead in the group. Allport might have observed, in inspecting his Table II, (1) that 9 of the subjects who reported more associations when working in the group actually *had more trials* when grouped, and (2) that if 11.4 A-trials produced 60.3 associations, the 13.5 T-trials

might have been expected (without "social facilitation") to produce above 70 associations—not (as reported) 63.6.

We have treated in a similar way Allport's quantitative results from his Exper. 3, 4, and 6 (figures are wanting for Exper. 5). All four experiments are summarized in our Table II. In all of

TABLE II

SUMMARY OF EXPERIMENTS 2–4, 6

Experiment	Aver. No. Trials per Subj.		Aver. No. Assoc. per Trial		M.V. from Aver. of Aver. No. Assoc. per Trial		No. Subj. with Excess Assoc. Favorable to:		No. Subj. with Greater (or Equal) No. Trials for A. and T.		
	A.	T.	A.	T.	A.	T.	A.	T.	A.	T.	Equal
2....	11.4	13.5	5.2	4.6	1.65	1.75	8	7	4	10	1
3....	5.5	5.7	21.02	20.35	5.28	4.35	6	8	2	5	7
4....	8.	11.	3.68	3.13	1.76	1.91	7	1	1	6	1
6....	16.6	17.7	0.89	0.62	0.79	0.27	4	5	3	6	0

them the average number of associations[6] for the average trial was *greater in isolation* than in the group; and, again, 25 subjects exceeded in this condition, where only 21 subjects turned out more associations (on the average per trial) when grouped. This evidence certainly does not appear to sustain our author's inferences upon "social facilitation."

Additional experiments (reported in the *Social Psychology*, pp. 274–278), directed toward the estimation of the degree of pleasantness and unpleasantness and of light and heavy weights, are thought by the author to "establish" the fact of "shunning extremes and expressing more moderate estimates when in the presence of other judges." Wanting individual records, measures of variability, adequate reports of the subjects, and the precise experimental conditions, it is impossible to decide whether this wide induction upon "social conformity" is really supported by the evidence at hand.[7]

[6] The unit-association (the basis of A's results) is, as everyone knows, a highly variable and uncertain factor. Whether it is sufficiently fixed to be indicative in this connection we doubt.

[7] It is not clear how one continuous scale of affective degrees could be used for both pleasantness and unpleasantness. Furthermore, we are not told whether the notable absorption of odors by clothing, the admixture of smell-stimulus and body-odor, and the more intensive air currents of the group were properly taken into account.

We have one further point in criticism to offer. Were Allport's deductions from the figures justified, we might still complain at his conception of "social response" and the use which he makes of it in this context. One may of course arbitrarily assume sociality among physically associated individuals who stand or sit within sight and hearing of each other, and one may further assume that the sociality rests upon, and is inherent in, the perceptions and movements incident to such a physical conjunction. One would then be inclined to deny—as Allport seems to deny—sociality to the solitary person and always and only to find it in the group. But such assumptions would overlook, at the very least, two important conditions of socialization. In the first place, a person is as obviously socialized by the sight of a personal gift, a coat thrown into the gutter, or a jimmied pantry window, as he is by the jostling throng or the call of a street-peddler; and, secondly, a personal relation considered in memory or imagination, that is to say without physical presence, is quite as *social* a relation as propinquity could possibly be. Social significances and references can enter, then, into (1) the perception of lifeless objects and into (2) absent persons as really as by way of the sight and sound of a physical associate. And if they can so "enter", then why not conceivably "facilitate" or "inhibit" performance? It seems to be only a determination to adhere to the naïve conception of sociality as depending upon the immediate give-and-take of congregated animals, upon *stimulus-and-response,* in the behavioristic phrase, that leads to the problem of facilitation as outlined by our author. In experiments of our own, where we have sought in vain for a general facilitating effect of sociality, it has appeared that sociality can never be imposed by, or assumed from, "presence" or "absence," but that it must be authenticated by the socialized person himself.

QUALITATIVE RESEMBLANCE AMONG ODORS

By Madison Bentley

However they may have fallen short of a definitive classification of odors, the experimental studies since Henning have at least made it clear that the investigations must now be carried on with a refined technique, in large numbers, and with carefully and specially trained observers. The vast number of exciting substances, the varied appeal to the receptors, olfactory and non-olfactory, and the complex qualitative relationships among the odors themselves combine to make the task exceedingly difficult. Henning's preliminary observations and his proposal of a classificatory scheme have at once aroused new interest and suggested new modes of attack upon an ancient psychological problem.[1]

Our own small contribution lies in the direction of additional controls and safeguards and in the proposal of slight methodical changes. We have taken fifty odorous substances (thirty-three of them chemically definable) and we have tried to determine the qualitative relationships of these fifty to three standards, chosen somewhat arbitrarily to stand near the spicy, fragrant and ethereal "corners" of Henning's figure. Our standards were S = nutmeg, F = heliotrope and E = lemon oil. The E was, unfortunately, a complex and undetermined substance.[2] Everyone who has worked with the problem knows the difficulties of discovering satisfactory representatives of these landmarks upon the proposed smell-figure. With care the three standard odors were, however, kept sensible constant through the experiments. The exact place which they may occupy in Henning's system is of no particular consequence to our own problem.

[1] H. Henning, *Das Geruch*, 2nd ed., 1924; F. L. Dimmick, *Amer. J. Psychol.*, 1922, 33, 423–425; M. K. Macdonald, *ibid.*, 1922, 33, 535–553.

[2] The standards and the unknowns were chosen without the writer's suggestion that he might observe them "without knowledge." The arrangements were made by Miss L. M. Hatfield, who served as experimenter throughout the research with objectivity, carefulness and high intelligence.

Care in presenting the substances, in respiration, and in the confusion of smells with other qualities has been much dwelt upon in recent studies; although the standardization of these matters has not gone very far. A long series of preliminary experiments directed us toward the control of olfactory intensities (concentration and distance of the exciting substance) and toward the elimination of disturbing circumstances. Our vapor current was dirhinic. E carefully governed the time of exposure by means of a silent metronome. Rate, order, interval, and other sources of variable error were provided for. O sat near an elevator shaft toward which a small motor-fan, run at moderate speed, was directed; but he sat just outside the main current of air. By leaning slightly forward he could bring his face into the current and so improve ventilation. E's manipulation of the small black, glass-stoppered containers went on within the air current between O and the shaft. The procedure was strictly without knowledge, save when the writer observed, and then O's information only included the general setting and not the number, order or names of the stimuli. Excitation of the blindfolded observer was usually restricted to one inhalation, which was taken under the following general instruction.

"Make yourself as comfortable as possible and try to maintain an attitude of affective indifference throughout the experiment. Before each stimulus is presented I shall say 'Ready,' and at the word 'Now' I shall present the stimulus. I want you to inhale deeply but steadily, being careful not to sniff, and then to exhale rather suddenly. Attend carefully to the olfactory quality, and to that alone. Report if you notice any temporal change in the quality. Be careful to base your comparisons upon quality, and not on pungency, affectiveness or any verbalizations which might be attached to the stimulus."

We laid great stress upon comfort, bodily relaxation, complete divorce of O from the setting and the manipulations, absolute quiet and absence of chatter with E, strict training of E in a monotonous and nonsuggestive set of signals, prompt exhalation of the breath, and *a set for olfactory quality only.* It is now apparent that contaminations from successive excitations easily damage olfactory results. Our sudden exhalation and the immediate removal of the vaporous air by currents greatly reduced the difficulty and at the same time notably prevented intensive reduction under exhaustion.

Our chief method employed three successive odors, the first and second of which represented the "standards." The special instruction here, given at the beginning of each series of ten, was as follows:

"I am going to give you three olfactory qualities in the order one, two, three. Report whether three is more like one or two, and give the degree of resemblance in terms of 'small,' 'medium' and 'great.' If it is like neither, report 'unrelated.' Give also the degree of your assurance in terms of 'low,' 'moderate' and 'high.'"

The main inspection concerned the three qualities, with the *Aufgabe* to report relative likeness. The interpretation of the reports, especially the discovery of faulty and changing criteria, was aided by the commentaries of O. Albeit, the better the

TABLE I

List of Odorous Substances

1. n-Caproic Acid
2. n-Heptoic Acid
3. n-Butyl Alcohol
4. Phenyl Ethyl Alcohol
5. Chloral Hydrate
6. Menthol
7. Quinoline
8. Dimethyl Aniline
9. n-Butyl Acetate
10. Diethyl Diethyl-malonate
11. Ethyl n-Valerate
12. Ethyl Cinnamate
13. Methyl Salicylate
14. Anethole
15. Anisol
16. o-Bromotoluene
17. p-Cymene
18. p-Dichlorobenzene
19. Essence of Wintergreen *
20. n-Propyl Bromide
21. Acetophenone
22. Phenol
23. Eugenol
24. Camphor
25. Pinene
26. Essence of Sassafras *
27. Di-n-butyl Amine
28. Vanillin
29. Coumarin
30. Geraniol
31. Benzyl Chloride
32. Ethyl Hexahydrobenzoate
33. Oil of Cloves *
34. p-Toluene Sulfonylchloride
35. Essence of Peppermint *
36. n-Ethyl Acetanilide
37. Iodine *
38. Listerine *
39. Fish Oil *
40. Strong Cheese *
41. Citral
42. Mutton Tallow *
43. Nitrobenzene *
44. Oil of Mace *
45. Caryophyllorum *
46. Syrup of Sarsaparilla *
47. Oil of Juniper *
48. Sulphuric Ether *
49. Oil of Terebinth *
50. Naphthalene *

* Chemically undetermined.

arrangements of E and the more reliable the O, the fewer the comments! All reports of "low assurance" were later repeated. At the end of each series O made an estimate of its reliability. Where this estimate was low and where other comments gave evidence of unfavorable conditions for observation, the entire

series was repeated at another time. Three series of 10 reports each filled the observation-hour. Each of the fifty unknowns was presented at least once with each possible combination of two of the "standards"; F = fragrant, E = ethereal and S = spicy. These combinations were FE, ES and SF, given in the two temporal orders.

The experiment was carried through with three junior students who had had, at the least, the general laboratory training. The main difficulty with such Os is the confusion of the object-smell with the olfactory quality (*Gegebenheitsgeruch*). Despite our caveats and despite the want of visual and other outside knowledge, the commentaries make it evident that (1) substances were at times "recognized" and then compared as objects and (2) extraneous qualities (pressure, pain, coldness, and the like) were used as criteria. The results accord fairly well with like results of other observers. Where a comparison with Henning (H), Dimmick (D) and Macdonald (M) is possible, we have set it down for whatever it may be worth. (Subscripts in the D-results indicate the number of placements.) Our Os agreed fairly well upon unknowns 8, 10, 14, 16, 19, 23, 28, 33, 41 and 44. The origin of some discrepancies appears in the commentaries; but other disagreements stand unexplained.

No. 6. Menthol. H(LS), D($S_{10}E_6$), M(near the center of the FESR face of the prism). All of our Os placed this on the SE side.

No.14. Anethole. H(S), M(S standard). Our Os placed this most often at S, but they also found a primary relationship with F and a slight resemblance to E.

No. 15. Anisol. H and M(S). Our Os found it unrelated.

No. 24. Camphor. H(R), D($R_{10}S_6E_1$). Our Os placed this on the SF side, weighted slightly on F.

No. 25. Pinene. H(R), M(R and near the center of the FESR face). Our reports are contradictory. Each O placed it at least once near each of the three corners, but also "no relation" was reported.

No. 26. Essence of Sassafras. H(SR), D($S_{10}E_8F_4$). This seems to be a triplex odor. It was placed eight times near S, six near E, and eight times near F.

No. 28. Vapillin. H(FS), D(E), M(F). All Os placed this on the FS side, nearer F.

No. 29. Coumarin. H(FS), D($F_{10}E_7S_4R_1$). Placements near F, near E, and near S.

No. 30. Geraniol. H(FE), M(ES near the middle). *R*, a simplex F-odor; *W* and *I*, resemblance to all three standards.

No. 33. Oil of Cloves. H(FS), D($S_{12}E_4F_8$). All Os; nearer S than any other corner, though small resemblance to F and E.

No. 35. Essence of Peppermint. D($S_{10}R_6E_6F_3$). Resemblance as reported by our Os is greatest toward S, with relationship also to E and S.

No. 41. Citral. H(E), M(E standard). Our Os; resemblance to S, E and F.

No. 47. Oil of Juniper. H(R), D($R_{11}E_4S_3$). *R*, a triplex odor; *W*, consistently near E, and *I* at S.

No. 48. Sulphuric Ether. H places most ethers at E. Our Os placed Sulphuric Ether most often at E; but *W* also with F; *R* reported it a simplex E, and *I* as wholly unrelated.

An entire set was also carried out, under the method of " three stimuli," with the writer serving as observer. The greatest care was taken to derive the report *immediately* from the olfactory qualities. Frequently the standards themselves were not recognized and verbalized because of the formal- and self-instruction to take each quality as simply " this " and so to set the relative degrees of relationship. The results for observer B are tabulated in Table II. In the vertical columns are recorded the placing of the fifty unknowns with respect to FE, ES and SF. Capital letters in the table stand for high relationship, small letters for low, coefficients for the number of times reported, and U for " unrelated."

Only two unknowns (25, 49) appear as related only to the ethereal standard; one (No. 36) to the S-standard only, and none only to F. Twelve (in Class IV) stood related to E and S only, and three others showed only a trace of F-ness. Only four (Nos. 9, 19, 29, 39 in Class V) excluded the S-relationship entirely. The largest single class (VII) contains 16 odors which display a general relationship to E, F and S; and 6 (Class VIII) stand almost entirely unrelated to all three. The remoter relation to F, for our whole series, is shown in the total entries; F = 54, S = 87, E = 90. Nevertheless, our Os (as well as Dimmick and Macdonald) noticed the small qualitative distance separating F and E. Where our relationships were equivocal we followed the set with a second lot (*two-stimuli* set) in which a given unknown was compared (for close, remote, or no relation) with *one standard at a time*. These results are put down in the three narrow columns, under the headings " E," " S " and " F." They confirm very nicely the relationships of the unknowns and

Table II

Number	F	E							S							F
Class I Related to E only																
25				2E			e				2U					U
49			e				e									
Class II: Related to S only																
36		U							S				S			
Class III: Related to F only																
					None											
Class IV Related chiefly to E & S																
2		2U	e		e		e	2U	2S	S	3U		S			
3			2e		U	E		U	S		S		S			
12			e			E	2e				S		S			
13				E	2E	2E	e		S		S		S			
30			e		E		e		S		S	S	S			
32				E		E					S		S			
40			2e					U	S	2S	SU			3U		
41			e			E					2S	S				
43				E	e					S		S				
48				E		2E					SS	S				
21		U		E	E					S			S			
15				E	2e				S			S	2S	U	2f	
18				E	E				S	S	U		S	U	2f	
38				E	e					2S			S		f	
22			e		U		e						S			
Class V: Related chiefly to E & F																
29				E		2E									f	F
39			e		2e	E	2e								f	F
9	F	f			2E	2E										F
11				E		E	2e		S						f	F
19				F		2E									F	2F
Class VI: Related chiefly S & F																
33	F		U		•					S		S				2F
37	F								S		S	S				UF
24		f							S	S					2f	F
14	F						e			2S		2S			F	f
42			2U	e				U			S	S			f	F	2f
Class VII: Related Equally to E S & F																
4					E						S	S			3f		
5				e	E		E	e	U	S	S				f		2f
6	F	f				e	E	e			2S		2S		f	F	3F
7		3f	U		E	E	2E		U		S	S	S		f		
17			U	e	E		E	e		S	S	S		U	2f	F	2f
10				e	3E			e			S	S	S			F	F
20		f	U	e		VE		2e		S	S	S	S			f	
23				e		e					2S		S			F	F
28		f					E			S	S	S				F	
35	F	f	U			E	E					S	2S			F	
44					E	E					S	S				F	FU
45				e	2E						S					F	
46	F					2E	E					S	S				F
47		f				E		e				S		2S			2f
1		f		2e				2e				S		S			
26					E	E					S	S				F	F
Class VIII: Slightly related																
8			U					e	U						2U		
16			U							2S		2U		2S	U	f	U
27		f	U					3e	U				S	S	U		
31			U	e								U		S			
34			U					e	U			S		S			U
50			U	2e			E	e						S	U	f	

substantiate the less-tried method of comparing two *qualitative-relations* (three-stimuli method).[3]

It is impossible to say, as it has been in other investigations, how many of the unknowns were single qualities. It is impossible, *e.g.*, to say whether *geraniol* (No. 30) resembles E and S as orange resembles a given red and a given yellow or as a tonal fusion resembles (by way of its several components) two neighboring tones. Before the olfactory qualities can be ordered, more knowledge of the simple items of olfactory experience must be acquired. Besides smell compounds, we have to contend with the interjection of warmth, cold, taste, pressure, tickle and pain. The writer was astonished to find what a large contribution to "smell" is made by the simple pressures of inhalation. He discovered that when a weak inhalation (without the intent to "smell in" a substance) cancelled these pressures, a single odorous quality was *totally unlocalized*. Regarded as a moment in experience, it was no more in-the-nose than it was in-the-hand or in-the-outside-air. This result has since been strikingly confirmed by Skramlik,[4] who found that the accessory qualities (which he severally disposes of in the nasal cavity) supplied the localization. He gives a long list (p. 104) of chemical substances which are without accessory effect and which he finds to be unlocalizable. Both our experiments and Skramlik's suggest that this lack of reference to the sense organ in "pure smells" may be of first-rate importance in setting qualitative relations.

Our small experiment suggests that the hope of discovering order and arrangement among the olfactory qualities will be remote until the simple qualities are directly compared without admixture of non-olfactory elements. The task of the observer

[3] After our experiments were completed (1924) appeared Findley's use of the three-stimulus method (*Amer. J. Psychol.*, 1924, 35, 436–445). In his case the unknown was presented *between* the two standards where, in our experiments, it was given *after* them. Only one standard (E = ethereal) and none of comparison-stimuli were common to the two investigations. Findley's Os gave but little consistency in their degrees of relationship, and he concludes that "it is thus impossible to verify Henning's qualitative theory with any great degree of precision" (445).

[4] E. v. Skramlik, Ueber die Lokalization der Empfindungen bei den niederen Sinnen. *Zsch. f. Sinnesphysiol.*, 1925, 56, 69–140.

should be further simplified; his instructions should be more specific and constant, and the criteria used in his comparison should be standardized with greater care. With adequately trained observers the direct comparison of similarity-differences is feasible, and it promises, under rigid control, to advance our knowledge of the olfactory system.